

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 17, 2000, 04:39:02 ; Search time 1878.33 seconds
(without alignments)
-1517.972 Million cell updates/sec

Title: US-09-151-189-1
Perfect score: 2931
Sequence: 1 CGCGGCAAGCCTTGGAAGA.....aaatggcgatagattatccc 2931

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0
Searched: 882769 seqs, -486395729 residues

Total number of hits satisfying chosen parameters: 1765538

Minimum DB seq length: 0

Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: GenEmbl:

- 1: gb_ba1.*
- 2: gb_ba2.*
- 3: gb_ba3.*
- 4: gb_ov.*
- 5: gb_pat.*
- 6: gb_p1.*
- 7: gb_p2.*
- 8: gb_p3.*
- 9: gb_p4.*
- 10: gb_p5.*
- 11: gb_p6.*
- 12: gb_p7.*
- 13: gb_p8.*
- 14: gb_p9.*
- 15: gb_p10.*
- 16: gb_p11.*
- 17: gb_p12.*
- 18: gb_p13.*
- 19: gb_p14.*
- 20: gb_p15.*
- 21: gb_p16.*
- 22: gb_p17.*
- 23: gb_p18.*
- 24: gb_p19.*
- 25: gb_p20.*
- 26: gb_p21.*
- 27: gb_p22.*
- 28: gb_p23.*
- 29: gb_p24.*
- 30: gb_p25.*
- 31: gb_p26.*
- 32: gb_p27.*
- 33: gb_p28.*
- 34: gb_p29.*
- 35: gb_p30.*
- 36: gb_p31.*
- 37: gb_p32.*
- 38: gb_p33.*
- 39: gb_p34.*
- 40: gb_p35.*
- 41: gb_p36.*
- 42: gb_p37.*
- 43: gb_p38.*
- 44: gb_p39.*

- 45: gb_p40.*
- 46: gb_p41.*
- 47: gb_p42.*
- 48: gb_p43.*
- 49: gb_p44.*
- 50: gb_p45.*
- 51: gb_p46.*
- 52: gb_p47.*
- 53: gb_p48.*
- 54: gb_p49.*
- 55: gb_p50.*
- 56: gb_p51.*
- 57: gb_p52.*
- 58: gb_p53.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Match	Length	DB	ID	Description
1	2931	100.0	2931	8	AF053411	Fucus dis
2	64.6	2.2	1794	5	E17199	E17199 Corallina p
3	64.6	2.2	2035	7	D87657	D87657 Corallina p
4	63.8	2.2	2039	7	D87658	D87658 Corallina p
5	62.8	2.1	1791	5	E17200	E17200 Corallina p
6	41.4	1.4	28043	42	AC014190	AC014190 Drosophila
7	41.4	1.4	75282	32	DNBR37M19	DNBR37M19 Drosophila
8	38.4	1.3	38400	1	SCCH2	SCCH2 Streptomy
9	38.6	1.3	42816	2	SCCH2	SCCH2 Streptomy
10	38.2	1.3	188975	52	AC022255	AC022255 Homo sapi
11	37.6	1.3	72683	55	AC023554	AC023554 Homo sapi
12	37.6	1.3	107677	10	HS40021	HS40021 Human DNA s
13	37.2	1.3	151496	32	AF000648	AF000648 Homo sapi
14	37	1.3	156457	45	AC016484	AC016484 Homo sapi
15	37	1.3	200965	53	AC020748	AC020748 Homo sapi
16	36.6	1.2	2317	8	OSU72250	OSU72250 Oryza sativ
17	36.6	1.2	63826	45	AC021539	AC021539 Homo sapi
18	36.6	1.2	79283	45	AC008465	AC008465 Homo sapi
19	36.6	1.2	100808	10	HS94616	HS94616 Human DNA s
20	36.6	1.2	123358	8	AC002396	AC002396 Arabidops
21	36.6	1.2	161449	42	AC012139	AC012139 Homo sapi
22	36.6	1.2	191904	40	AC005699	AC005699 Homo sapi
23	36.4	1.2	711	8	MDU77952	MDU77952 Malus domes
24	36.4	1.2	67568	44	AC020378	AC020378 Drosophila
25	36.4	1.2	186323	12	AC006491	AC006491 Drosophila
26	36.4	1.2	188914	33	AL138710	AL138710 Homo sapi
27	36.4	1.2	205328	33	AL137219	AL137219 Homo sapi
28	36.2	1.2	3162	12	MUSVIA3	MUSVIA3 House mRNa
29	36	1.2	1256	12	MUSCIRPB	MUSCIRPB Mus musculu
30	36	1.2	7218	8	I66484	I66484 Sequence 14
31	36	1.2	42603	2	AF188287	AF188287 Stigmatal
32	36	1.2	289420	57	AC012153	AC012153 Homo sapi
33	35.8	1.2	797	11	HSU80998	HSU80998 Human basic
34	35.8	1.2	2298	1	MSHISC	MSHISC M.smeagmatis
35	35.8	1.2	78359	44	AC018941	AC018941 Homo sapi
36	35.8	1.2	188205	40	AC003986	AC003986 Homo sapi
37	35.6	1.2	161506	54	AC016398	AC016398 Homo sapi
38	35.4	1.2	1280	7	ABSPR1	ABSPR1 Agarius bi
39	35.4	1.2	4557	1	HSCHEA	HSCHEA H.salinariu
40	35.4	1.2	98734	32	PFMAL1P2	PFMAL1P2 Plasmodiu
41	35.4	1.2	137466	43	AC013597	AC013597 Homo sapi
42	35.2	1.2	14477	34	DMU02279	DMU02279 Drosophila
43	35	1.2	931	35	AF017665	AF017665 Manduca s
44	35	1.2	5037	1	MSASDASK	MSASDASK M.smeagmatis
45	35	1.2	58195	55	AC023622	AC023622 Mus muscu

ALIGNMENTS

41491 41590: gap of unknown length
41591 43322: contig of 1732 in length
43323 43422: gap of unknown length
43423 44000: contig of 578 in length
44001 44100: gap of unknown length
44101 44815: contig of 715 in length
44816 44915: gap of unknown length
44916 45445: contig of 530 in length
45446 45545: gap of unknown length
45546 47856: contig of 2311 in length
47857 48320: contig of 874 in length
48321 48930: gap of unknown length
48931 49449: contig of 519 in length
49450 49578: gap of unknown length
49579 50078: contig of 429 in length
50080 50853: gap of unknown length
50854 50953: gap of unknown length
50954 51481: contig of 528 in length
51482 51581: gap of unknown length
51582 51900: contig of 319 in length
51901 52000: gap of unknown length
52001 52172: contig of 172 in length
52173 52272: gap of unknown length
52273 52750: contig of 478 in length
52751 52850: gap of unknown length
52851 53511: contig of 661 in length
53512 53611: gap of unknown length
53612 54312: contig of 701 in length
54313 54412: gap of unknown length
54413 54764: contig of 352 in length
54765 54864: gap of unknown length
54865 55933: contig of 1069 in length
55934 56033: gap of unknown length
56034 57299: contig of 1286 in length
57300 57999: gap of unknown length
58000 58424: contig of 1025 in length
58425 58524: gap of unknown length
58525 60083: contig of 1559 in length
60084 61374: gap of unknown length
61375 61474: gap of unknown length
61475 62005: contig of 531 in length
62006 62105: gap of unknown length
62106 62559: contig of 554 in length
62560 62759: gap of unknown length
62760 63255: contig of 496 in length
63256 63355: gap of unknown length
63356 63909: contig of 554 in length
63910 64009: gap of unknown length
64010 66317: contig of 2308 in length
66318 66417: gap of unknown length
66418 67225: contig of 808 in length
67226 67325: gap of unknown length
67326 69889: contig of 2564 in length
69890 70539: gap of unknown length
70540 70639: gap of unknown length
70640 71142: contig of 503 in length
71143 71242: gap of unknown length
71243 71792: contig of 550 in length
71793 71892: gap of unknown length
71893 74273: contig of 2381 in length
74274 74373: gap of unknown length
74374 75282: contig of 909 in length.

* NOTE: This is a 'working draft' sequence.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

FEATURES
Source

Location/Qualifiers
1..75282
/organism="Drosophila melanogaster"

/strain="v: cm bw sp"
/db_xref="taxon:7227"
/chromosome="X"
/clone="BACR37M19"
/clone_lib="RPCI-98 (Roswell Park Cancer Institute)
/prospecies="Drosophila melanogaster BAC library, partial SCORI in
pBAC3.6
/map="10E".
BASE COUNT: 19878 a 14540 c 14571 g 19191 t 7102 others
ORIGIN
Query Match 1.4%; Score 41.4; DB 32; Length 75282;
Best Local Similarity 54.2%; Pred. No. 3.9;
Matches 84; Conservative 0; Mismatches 71; Indels 0; Gaps 0;
QY 1088 cgtggaatggagggtctcccaaatctggagcgcgtgtcgatag99tccgat99tac99t 1147
DB 19436 CGATGGACTTGGCGGATGCCAATCTGCTCCGGGTGACCGCATCCGATCTGGCGA 19495
QY 1148 ggaccgtttctccagctcttcggagcgaccttctgtgtgttgaaaggggaccttctgt 1207
DB 19496 AGAGTTTGACGCACATGCTCGGACGTCGAGCGCTGTTGTGCTCAGGAAGTCAGATGCT 19555
QY 1208 ctctcagctctcgtagaacagcttcacacatcgacg 1242
DB 19556 TGCCCATCTGCTCGAGAGCGGCATATCCGTCGGCG 19590
RESULT 8
SC4H2 SC4H2 38400 bp DNA BCT 06-APR-1998
LOCUS Streptomyces coelicolor cosmid 4H2.
DEFINITION Streptomyces coelicolor cosmid 4H2.
ACCESSION AL022268
VERSION AL022268.1 GI:3036873
KEYWORDS aminotransferase; ATP-dependent helicase; dapF; diaminopimelate
epimerase; hydrolase; kinase/phosphohydrolase; lexA; nrdJ; secreted
hydrolase; SOS regulatory protein LexA; transcriptional regulator;
transmembrane efflux protein; tRNA
delta(2)-isopentenylpyrophosphate transferase; two-component
system; zinc metalloprotease.
SOURCE Streptomyces coelicolor A3(2).
ORGANISM Streptomyces coelicolor A3(2).
REFERENCE 1. (bases 1 to 38400)
Actinomycetales; Streptomycinae; Streptomycetaceae; Streptomyces.
AUTHORS Seeger, K.J. and Harris, D.
JOURNAL Unpublished
REFERENCE 2. (bases 1 to 38400)
Parkhill, J., Barrell, B.G. and Randal, M.A.
JOURNAL Direct Submission
TITLE Submitted (24-MAR-1998) Streptomyces coelicolor sequencing project.
ABSTRACT Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof.
David A. Hopwood, [3] John Innes Centre, Norwich Research Park,
Colney, Norwich, Norfolk NR4 7UR, UK
REFERENCE 3. (bases 1 to 38400)
Redenbach, M., Kieser, H.M., Denapate, D., Eichner, A., Cullum, J.,
Kinashi, R. and Hopwood, D.A.
AUTHORS A set of ordered cosmids and a detailed genetic and physical map
for the 8 Mb Streptomyces coelicolor A3(2) chromosome
JOURNAL Mol. Microbiol. 21 (1), 77-96 (1996)
MEDLINE 97000351
COMMENT On Apr 8, 1998 this sequence version replaced gi:2995287.
Notes:
Streptomyces coelicolor sequencing at the Sanger Centre is funded
by the BBSRC.
Details of S. coelicolor sequencing at the Sanger Centre are
available on the World Wide Web.
(URL: <http://www.sanger.ac.uk/Projects/S.coelicolor/>) CDS are
numbered using the following system eg SCB7.0lc, SC (S.
coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary
strand).

The more significant matches with motifs in the PROSIVE database are also included but some of these may be fortuitous. The length in codons is given for each CDS.

Usually the highest scoring match found by fasta -o is given for CDS which show significant similarity to other CDS in the database. The position of possible ribosome binding site sequences are given where these have been used to deduce the initiation codon. Gene prediction is based on positional base preference in codons using a specially developed hidden Markov Model (Krogh et al., Nucleic Acids Research, 22(22):4768-4778(1994)) and the frameplot program of Elblb et al., Gene 30:157-166(1984) as implemented at <http://www.nih.gov/jp/jun/cgi-bin/frameplot.pl>.

CADTION: We may not have predicted the correct initiation codon. Where possible we choose an initiation codon (atg, gta, ttg or att) which is preceded by an upstream ribosome binding site sequence (optimally 5-1bp before the initiation codon). If this cannot be identified we choose the most upstream initiation codon.

UPSTREAM: This sequence MAY NOT be the entire insert of the sequenced clone. It may be shorter because we only sequence overlapping sections once, or longer, because we arrange for a small overlap between neighbouring submissions. Cosmid 4H2 lies on the AseI B genomic fragment between St4H8 and St5H8.

Location/Qualifiers

1. 38400

misc_feature

gene

CDS

RBS

gene

CD5

gene

505

gene

CDS

RBS

RES

RES

σ₀

33

AAALTLPLPLGAGVELLVDGGLGTGAAGAGTGRKRRALLIGAEIHFEPAPGG
GTDVRLHVPAPGDRCADRTGDSF

gene

4687..5346

CDS

/gene="SC4H2.06"

4687..5346

/gene="SC4H2.06"

/note="SC4H2.06, probable two-component transcriptional
regulator, len: 219 aa; similar to many e.g. PRG5, BACS5
P13800 transcriptional regulatory protein degg (229 aa),
fasta scores: opt: 343 z-score: 736.1 E(0.07, 37.14)
identity in 224 aa overlap. Contains PS00622 Bacterial
regulatory proteins, luxR family signature, P1am match to

Query Match 1.3% Score 39.4; DB 1; Length 38400;

Best Local Similarity 44.2%; Pred. No. 12; Mismatches 206; Indels 0; Gaps 0;
Matches 163; Conservative 0;

QY 1150 accgttctccagctcttcgagcgaccttgggtgtgaaacggggccctttgtct 1209

DB 13012 ACCGGCTCGCGGCTACCTGATGCTCGGCAACTGGCGGAGAGACCAACCTGACCG 13071

QY 1210 ctgagctgctgagcagcttcacatcgacgctattacggtcgaaacgagcagagaga 1269

DB 13072 CTTGCTCTCCACGACCGCTCGAAGACACCGAGGTGAGCTCGATCAGTTCGCGAGC 13131

QY 1270 cattgcgccgactgaactatattgctgcttttgcgaatggtgaacattcagaatg 1329

DB 13132 AGTTCGCGCGGAGTTCCTACTCTGTCGACGCGCTCAAGCTGGAGAGTCTGACT 13191

QY 1330 gtgagcccgccgagcgcccgagaggttagacgaagagctgttttattcgttaacgcc 1389

DB 13192 ACGGCGCGCGCGGAGCGCGGAGACCTTCGCAAGAGTCTCTGCCACCGCAAGAG 13251

QY 1390 gcaactgcccagggctctctctgtagaatatcatcaacgaagcttatcggggtctc 1449

DB 13252 TCCGGTGATGTCATCAATCTCGCGACCGCTGCAACATGCGCACCTCTGCTGTGCA 13311

QY 1450 ttatctactgagctggagccttcagcccgcccgatcatcaacggttcattcatcagaca 1509

DB 13312 TGCCGCGGAGAGAGGAGGAGCAATCGCCAAAGGTACCGCGGAGTCTCTGCTGTGCG 13371

QY 1510 gtgatggc 1518

DB 13372 CCGAAGCGC 13380

RESULT 9

SCC88/c

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

MEDLINE

JOURNAL

AUTHORS

REFERENCE

AUTHORS

TITLE

SCC88 42816 bp DNA BCT 15-FEB-2000

Streptomyces coelicolor cosmid C88.

AL139298

505 ribosomal protein L21; 50S ribosomal protein L27;

dehydrogenase; fpgs: glycosyltransferase; integral membrane

protein; integrase; membrane protein; mrcB; mrcC; mrcD; ndk; obg;

hbp2; secreted protein; sfr; transactinase; valS.

Streptomyces coelicolor A3(2).

Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;

Actinomycetales; Streptomycetaceae; Streptomyces.

Redenbach, M., Kleser, H.M., Denapite, D., Eichner, A., Cullum, J.,

Kinashi, H. and Hopwood, D.A.

A set of ordered cosmids and a detailed genetic and physical map

for the 8 Mb Streptomyces coelicolor A3(2) chromosome

Mol. Microbiol. 21 (1), 77-96 (1996)

97000351

2 (bases 1 to 42816)

Unpublished

Unpublished

3 (bases 1 to 42816)

Cerdano, A.M., Parkhill, J., Barrell, B.G. and Rajandream, M.A.

Direct Submission

JOURNAL

COMMENT

Submitted (14-FEB-2000) Streptomyces coelicolor sequencing project,
Sanger Centre, Wellcome Trust Genome Campus, Hinxton, Cambridge
CB10 1SA E-mail: barrell@sanger.ac.uk Cosmids supplied by Prof.
David A. Hopwood, [3] John Innes Centre, Norwich Research Park,
Colney, Norwich, Norfolk NR4 7UH, UK

Notes:
Streptomyces coelicolor sequencing at the Sanger Centre is funded
by the BBSRC and Wellcome Genomics
Details of S. coelicolor sequencing at the Sanger Centre are
available on the World Wide Web.
(URL: <http://www.sanger.ac.uk/projects/S-coelicolor/>)

CDS are numbered using the following system eg SC7B7.01c. SC (S.
coelicolor), 7B7 (cosmid name), .01 (first CDS), c (complementary
strand).

The more significant matches with motifs in the PROSITE database
are also included but some of these may be fortuitous.
The length in codons is given for each CDS.

Usually the highest scoring match found by fasta -o is given for
CDS which show significant similarity to other CDS in the database.
The position of possible ribosome binding site sequences are given
where these have been used to deduce the initiation codon.

Gene prediction is based on positional base preference in codons
using a specially developed Hidden Markov Model (Krogh et al.,
Nucleic Acids Research, 22(22):4768-4778(1994)) and the FramePlot
program of Bibb et al., Gene 30:157-66(1984) as implemented at
<http://www.nih.gov/jp/>

Jun/Cri-bin/frameplot.pl. CAUTION: We may not have predicted the
correct initiation codon. Where possible we choose an initiation
codon (atg, gtg or (att)) which is preceded by an upstream
ribosome binding site sequence (optimally 5-13bp before the
initiation codon). If this cannot be identified we choose the most
upstream initiation codon.

IMPORTANT: this sequence MAY NOT be the entire insert of the
sequenced clone. It may be shorter because we only sequence
overlapping sections once, or longer, because we arrange for a
small overlap between neighbouring submissions.
Cosmid C88 lies between and overlaps with cosmids C123 and C80 on
the Aser-C genomic restriction fragment.

Location/Qualifiers

1. 42816

/organism="Streptomyces coelicolor A3(2)"

/strain="A3(2)"

/db_xref="taxon:100226"

/clone="cosmid C88"

1. 3409

/gene="SCC88.01"

1. 106

/gene="SCC88.01"

/note="nominal overlap with S. coelicolor cosmid StC123"

1. 3409

/partial

/gene="SCC88.01"

/note="Protein sequence is in conflict with the conceptual

translation: SCC88.01 possible glycosyltransferase, len:

>135 aa; N-terminal region similar to TR-Q9AC92

(EMBL:AF14632) Klebsiella pneumoniae glycosyltransferase,

329 aa; fasta scores: opt: 214 z-score: 245.4 E(0.3e-06);

21.3% identity in 263 aa overlap and C-terminal region

similar to SW:TAGF_BACSV (EMBL:X15200) Bacillus subtilis

telcholic acid biosynthesis protein F TagF, 746 aa; fasta

scores: opt: 1056 z-score: 1203.0 E(0.41e-5) identity

in 381 aa overlap"

/codon_start=2

/transl_table=11

/product="putative glycosyltransferase"

/protein_id="CA975371.1"

/db_xref="GI:6983732"

/translation="MRKAGAGARTGTGTRLAGSEVAVFVDSMDIPDVAFLAV

ESLDESGDFVGNVQHSKINQSPHRLAGAVRTHSRNKKILLTRACNV

FRSEFVRLDLPFQVLYEDVYLLPAQLAENVDIISEPTVYRLREGANPSITO

RRFEPKVRDRAAYESVSRFASPGSGAEKHAIDHALTGDLRIFLAVLPDGA

ETFAELRVINKVQVDPVRFAPAKIKMLVKKHALDVLNLSKRGDAVK

ISGLVKNVSYTVEANAAGLPKKARRIDPLRLHAPLQDIRWDCGRRLSLGSAWVO

Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,
 Subramanian, A., Talamas, J., Testaye, S., Theodore, J., Tirrell, A.,
 Travers, M., Triggillo, J., Vnsillie, H., Viel, R., Vo, A., Wilson, B.,
 Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zimmer, A. and
 Zody, M.

TITLE JOURNAL

COMMENT

Direct Submission
 Submitted (15-FEB-2000) Whitehead Institute/MIT Center for Genome
 Research, 320 Charles Street, Cambridge, MA 02141, USA
 All repeats were identified using RepeatMasker:
 Smith, A.F.A. & Green, P. (1996-1997)
 http://ftp.genome.washington.edu/RM/RepeatMasker.html

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence.submissions@genome.wi.mit.edu

----- Project Information

Center Project name: L6506

Center Clone name: 589_P23

* NOTE: This record contains 80 individual
 * sequencing reads that have not been assembled into
 * contigs. Runs of N are used to separate the reads
 * and the order in which they appear is completely
 * arbitrary. Low-pass sequence sampling is useful for
 * identifying clones that may be gene-rich and allows
 * overlap relationships among clones to be deduced.
 * However, it should not be assumed that this clone
 * will be sequenced to completion. In the event that
 * the record is updated, the accession number will
 * be preserved.

1
 911: contig of 911 bp in length
 912: gap of unknown length
 1833: contig of 922 bp in length
 1834: gap of unknown length
 2732: contig of 899 bp in length
 2733: gap of unknown length
 3630: contig of 898 bp in length
 3631: gap of unknown length
 4513: contig of 883 bp in length
 4514: gap of unknown length
 5412: contig of 899 bp in length
 5413: gap of unknown length
 6347: contig of 935 bp in length
 6348: gap of unknown length
 7231: contig of 884 bp in length
 7232: gap of unknown length
 8129: contig of 898 bp in length
 8130: gap of unknown length
 9052: contig of 923 bp in length
 9053: gap of unknown length
 9978: contig of 926 bp in length
 9979: gap of unknown length
 10924: contig of 946 bp in length
 10925: gap of unknown length
 11825: contig of 901 bp in length
 11826: gap of unknown length
 12706: contig of 881 bp in length
 12707: gap of unknown length
 13609: contig of 903 bp in length
 13610: gap of unknown length
 14511: contig of 902 bp in length
 14512: gap of unknown length
 15400: contig of 889 bp in length
 15401: gap of unknown length
 16316: contig of 916 bp in length
 16317: gap of unknown length
 17204: contig of 888 bp in length
 17205: gap of unknown length
 18114: contig of 910 bp in length
 18115: gap of unknown length
 19019: contig of 905 bp in length
 19020: gap of unknown length
 19444: contig of 925 bp in length
 19445: gap of unknown length
 20863: contig of 919 bp in length
 20864: gap of unknown length
 21773: contig of 910 bp in length
 21774: gap of unknown length
 22686: contig of 913 bp in length
 22687: gap of unknown length
 23617: contig of 931 bp in length
 23618: gap of unknown length
 24528: contig of 911 bp in length
 24529: gap of unknown length
 25425: contig of 897 bp in length
 25426: gap of unknown length
 26351: contig of 926 bp in length
 26352: gap of unknown length
 27273: contig of 922 bp in length
 27274: gap of unknown length
 28210: contig of 937 bp in length
 28211: gap of unknown length
 29139: contig of 929 bp in length
 29140: gap of unknown length
 30847: contig of 908 bp in length
 30848: gap of unknown length
 30908: contig of 861 bp in length
 30909: gap of unknown length
 31830: contig of 922 bp in length
 31831: gap of unknown length
 32733: contig of 903 bp in length
 32734: gap of unknown length
 33629: contig of 896 bp in length
 33630: gap of unknown length
 34512: contig of 883 bp in length
 34513: gap of unknown length
 35438: contig of 926 bp in length
 35439: gap of unknown length
 36308: contig of 870 bp in length
 36309: gap of unknown length
 37230: contig of 922 bp in length
 37231: gap of unknown length
 38157: contig of 927 bp in length
 38158: gap of unknown length
 39069: contig of 912 bp in length
 39070: gap of unknown length
 40024: contig of 955 bp in length
 40025: gap of unknown length
 40913: contig of 889 bp in length
 40914: gap of unknown length
 41801: contig of 888 bp in length
 41802: gap of unknown length
 42710: contig of 909 bp in length
 42711: gap of unknown length
 43661: contig of 951 bp in length
 43662: gap of unknown length
 44541: contig of 880 bp in length
 44542: gap of unknown length
 45435: contig of 894 bp in length
 45436: gap of unknown length
 46349: contig of 914 bp in length
 46350: gap of unknown length
 47269: contig of 920 bp in length
 47270: gap of unknown length
 48204: contig of 935 bp in length
 48205: gap of unknown length
 49085: contig of 881 bp in length
 49086: gap of unknown length
 49990: contig of 905 bp in length
 49991: gap of unknown length
 50894: contig of 904 bp in length
 50895: gap of unknown length
 51781: contig of 887 bp in length
 51782: gap of unknown length
 52679: contig of 898 bp in length

```

* 52680 53552: contig of 873 bp in length
* 53553 54466: contig of 914 bp in length
* 54467 55364: contig of 898 bp in length
* 55365 56295: contig of 931 bp in length
* 56296 57215: contig of 920 bp in length
* 57216 58155: contig of 940 bp in length
* 58156 59072: contig of 917 bp in length
* 59073 59968: contig of 896 bp in length
* 59969 60878: contig of 910 bp in length
* 60879 61785: contig of 907 bp in length
* 61786 62664: contig of 879 bp in length
*      gap of unknown length

Query Match      1.38; Score 37.67; DB 55; Length 72683;
Best Local Similarity 52.68; Pred. No. 43;
Matches 82; Conservative 0; Mismatches 74; Indels 0; Gaps 0;

QY 2774 ttgagtgacctgtaccagtggtgtaagacattttctctgtatggtatctgtaga 2833
Db 43942 TGAAGAAATGGAAGATTGGTGCTCTTAATATACATCTGTGATCTATTCCAA 44001

QY 2834 acagctagtggtgtttatcacagtagtctataaataagtaggtgtgataatgcac 2893
Db 44002 GGAGCTATTGTTATTACCAAGATTCAGTAATTCAGTAATTCAGTAATTCAGTA 44061

QY 2894 gttactatgaacacgcaaatggcgatagatttc 2929
Db 44062 GAGCTTCTAATACACCAATTCATCTGCTGGATATAC 44097

```

```

RESULT 12
LOCUS      HS440021      107677 bp.      DNA      PRI      23-NOV-1999
DEFINITION Human DNA sequence from PAC 440021 on chromosome X contains ESTs
and STS.
ACCESSION 284481
VERSION    284481.1 GI:1945153
KEYWORDS   human.
SOURCE     Homo sapiens
ORGANISM   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Ruminidae; Homo.
REFERENCE 1 (bases 1 to 107677)
AUTHORS    Pearce, A.
TITLE      Direct Submission
JOURNAL    Submitted (17-APR-1997) Sanger Centre, Hinxton, Cambridgeshire,
            CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk clone
            requests: clonerequest@sanger.ac.uk
COMMENT    On Apr 19, 1997 this sequence version replaced gi:1806016.
            de Jong P.J., enquires: http://bacpac.med.buffalo.edu/IMPORTANT/.
            This sequence is the entire insert of clone 440021. This sequence
            has been finished according to sequence map criteria as follows.
            An attempt is made to resolve all sequencing problems, such as
            compressions and repeats, but not necessarily within known
            annotated human repeat sequence elements (e.g. Alu). Where the
            sequence is ambiguous, there is an annotation using the 'unsure'
            feature key.
            The true left end of clone 440021 is at 1 in this sequence. The
            true right end of clone 440021 is at 107677.
            440021 is from the library APC13 constructed at the Roswell Park
            Cancer Institute by the group of Pieter de Jong. For further
            details see http://bacpac.med.buffalo.edu/VECTOR: PCIPAC2.

```

FEATURES

```

source
1. .107677
   /organism="Homo sapiens"
   /db_xref="taxon:9606"
   /chromosome="X"
   /map="X"
   /clone="RP3-440021"
   /clone_lib="RPCI-3"
   repeat_region
117..207
   /note="LIMM4 repeat: matches 961..1047 of consensus"
   repeat_region
331..1332
   /note="L1 repeat: matches 4392..5390 of consensus"
   repeat_region
1179..2209
   /note="LIMM3 repeat: matches 1..1051 of consensus"
   repeat_region
2917..3215
   /note="AluX repeat: matches 5..302 of consensus"
   repeat_region
3688..3877
   /note="MIR repeat: matches 247..41 of consensus"
   repeat_region
4993..5216
   /note="AluY repeat: matches 301..80 of consensus;
incomplete repeat"
   repeat_region
5239..5279
   /note="MIR28 repeat: matches 397..356 of consensus"
   repeat_region
5343..5789
   /note="MIR29A repeat: matches 395..8 of consensus"
   repeat_region
7550..7857
   /note="AluY repeat: matches 295..1 of consensus"
   repeat_region
9156..9324
   /note="L1 repeat: matches 4871..5035 of consensus"
   repeat_region
9652..9698
   /note="MIR2CB repeat: matches 501..455 of consensus"
   repeat_region
10208..10879
   /note="MIR2B repeat: matches 494..5 of consensus"
   repeat_region
11071..16448
   /note="L1 repeat: matches 9..5390 of consensus"
   repeat_region
16303..17189
   /note="LIM2 repeat: matches 1..891 of consensus"
   repeat_region
18419..18722
   /note="AluX repeat: matches 1..302 of consensus"
   repeat_region
18729..21555
   /note="L1 repeat: matches 2533..5384 of consensus"
   repeat_region
21419..21991
   /note="LIMC1 repeat: matches 6..549 of consensus"
   repeat_region
21995..22287
   /note="AluX repeat: matches 288..1 of consensus"
   repeat_region
22888..22817
   /note="LIMC1 repeat: matches 543..1079 of consensus"
   repeat_region
22859..23715
   /note="LIMB6 repeat: matches 64..922 of consensus"
   repeat_region
24116..25064
   /note="L1 repeat: matches 4623..3674 of consensus"
   repeat_region
25153..25452
   /note="AluX repeat: matches 302..1 of consensus"
   repeat_region
25455..26328
   /note="L1 repeat: matches 3686..2773 of consensus"
   repeat_region
27792..27829
   /note="L1 repeat: matches 100 & conserved"
   repeat_region
28635..29967
   /note="MIR repeat: matches 96..250 of consensus"
   repeat_region
30233..30345
   /note="LIMPA7 repeat: matches 579..892 of consensus"
   repeat_region
30949..31364
   /note="LIMB3 repeat: matches 896..482 of consensus"
   repeat_region
31347..31698
   /note="THEIC repeat: matches 2..371 of consensus"
   repeat_region
31700..32191
   /note="LIMM3 repeat: matches 501..1 of consensus"
   repeat_region
32043..33837
   /note="L1 repeat: matches 5390..3544 of consensus"
   repeat_region
33861..34173
   /note="AluX repeat: matches 301..1 of consensus"
   repeat_region
34500..34538
   /note="MIR2 repeat: matches 102..140 of consensus"

```

```

repeat_region 35783..36023
/Note="L1P4 repeat: matches 648..888 of consensus"
repeat_region 37225..37268
/Note="22 copies of 2 mer 93 & conserved"
repeat_region 38240..38439
/Note="MIR repeat: matches 222..32 of consensus"
repeat_region 39158..40100
/Note="L1M2 repeat: matches 1055..1 of consensus"
repeat_region 40054..40932
/Note="L1 repeat: matches 5390..4523 of consensus"
repeat_region 40973..41158
/Note="MIR repeat: matches 191..2 of consensus"
repeat_region 41204..41432
/Note="MER3 repeat: matches 209..6 of consensus"
repeat_region 41612..41707
/Note="MIR2 repeat: matches 53..146 of consensus"
repeat_region 43556..43854
/Note="AluSg repeat: matches 1..300 of consensus"
repeat_region 44857..44908
/Note="26 copies of 2 mer 92 & conserved"
repeat_region 44909..44985
/Note="MER34 repeat: matches 456..533 of consensus"
repeat_region 45260..46134
/Note="L1M3 repeat: matches 917..5 of consensus"
repeat_region 45989..46771
/Note="L1 repeat: matches 5390..4604 of consensus"
repeat_region 46768..47613
/Note="L1 repeat: matches 4574..3737 of consensus"
repeat_region 47622..47735
/Note="MER34 repeat: matches 521..410 of consensus"
repeat_region 47741..48041
/Note="AluSx repeat: matches 302..1 of consensus"
repeat_region 48042..48495
/Note="MER34 repeat: matches 431..8 of consensus"
repeat_region 52726..52789
/Note="L1M2 repeat: matches 1074..1011 of consensus"
repeat_region 52726..52834
/Note="L1M1 repeat: matches 957..839 of consensus"
repeat_region 52870..53229
/Note="THE1B repeat: matches 364..1 of consensus"
repeat_region 53233..53532
/Note="L1M1 repeat: matches 797..507 of consensus"
repeat_region 53785..53906
/Note="MIR2 repeat: matches 142..23 of consensus"
repeat_region 54257..54327
/Note="MER33 repeat: matches 1..71 of consensus"
repeat_region 54336..54631
/Note="L1M1 repeat: matches 66..358 of consensus"
repeat_region 54584..55367
/Note="L1M3 repeat: matches 256..1051 of consensus"
repeat_region 55374..55559
/Note="MER33 repeat: matches 53..246 of consensus"
repeat_region 55563..55856
/Note="AluDb repeat: matches 298..1 of consensus"
repeat_region 56799..57352
/Note="L1P413 repeat: matches 903..353 of consensus"
repeat_region 57364..59701
/Note="L1 repeat: matches 5297..3008 of consensus"
repeat_region 59701..60002
/Note="L1M2 repeat: matches 1..304 of consensus"
repeat_region 60033..60406
/Note="THE1B repeat: matches 364..1 of consensus"
repeat_region 60412..60627
/Note="L1M2 repeat: matches 329..553 of consensus"
repeat_region 60470..60627
/Note="L1M2 repeat: matches 295..453 of consensus"
repeat_region 60624..62683
/Note="L1 repeat: matches 3017..969 of consensus"
repeat_region 64460..64848
/Note="MIR repeat: matches 1..394 of consensus"
repeat_region 65511..65654
/Note="FLAN_A repeat: matches 1..133 of consensus"
repeat_region 66464..66765

```

```

/Note="AluSx repeat: matches 1..302 of consensus"
66789..66830
/Note="21 copies of 2 mer 81 & conserved"
67567..67918
/Note="L1M1 repeat: matches 19..365 of consensus"
68226..68349
/Note="MER34 repeat: matches 189..65 of consensus"
68351..68552
/Note="MIR repeat: matches 70..211 of consensus"
68633..68706
/Note="L1M2 repeat: matches 1055..983 of consensus"
68697..70419
/Note="L1 repeat: matches 3804..2022 of consensus"
70830..71387
/Note="L1 repeat: matches 2001..1428 of consensus"
<71389..>71710
/Note="match: 3' EST N58388 clone 248070"
72933..73493
/Note="L1M2 repeat: matches 553..1 of consensus"
73177..73493
/Note="L1M2 repeat: matches 309..1 of consensus"
74118..74451
/Note="MER2 repeat: matches 1..345 of consensus"
74959..74996
Query Match 1..38 Score 37.6; DB 10; Length 107677;
Best Local Similarity 52.64; Pred No. 46;
Matches 82; Conservative 0; Mismatches 74; Indels 0; Gaps 0;
Qy 2774 tgaagctactgtaccagtagtggtgaagacatatcttctgttattgtagctgtaga 2833
Db 2524 TGAAGAAATAGGAAGACTTTCGCTGCTAAATATACATCCCTTGCATCTATTTCAAA 2583
Qy 2834 acagctaggtgtgtttattacacagtagctataaataagtaggtgtgataatgcac 2893
Db 2584 GGAGCTATTGTTGTTTATACCAAGATTATCCAGTAATAAGAAAATCAAAATACTGT 2643
Qy 2894 ggaactatgaacacgaatggcgtagatattc 2929
Db 2644 GAGCTTCTAAACACCATATTCATCTGTGATATAC 2679
RESULT 13
AP000648 151496 bp DNA HTG 04-FEB-2000
LOCUS Homo sapiens chromosome 11 clone CMB9-76D16 map 11q14, WORKING
DEFINITION DRAFT SEQUENCE, 26 unordered pieces.
ACCESSION AP000648
VERSION 1 GI:5997525
KEYWORDS HTG; HGSC_PHASE1; HTGS_DRAFT
SOURCE Homo sapiens BDN, clone: CMB9-76D16.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1. (bases 1 to 151496)
AUTHORS Hattoni, M., Ishii, K., Toyoda, A., Taylor, T.D., Hong-Seog, P.,
Fujiyama, A., Yada, T., Totoki, Y., Watanabe, H. and Sakaki, Y.
Homo sapiens 151,496 genomic DNA of 11q14
Published Only in Database (1999) In press
JOURNAL
REFERENCE 2. (bases 1 to 151496)
AUTHORS Hattoni, M., Ishii, K., Toyoda, A., Taylor, T.D., Hong-Seog, P.,
Fujiyama, A., Yada, T., Totoki, Y., Watanabe, H. and Sakaki, Y.
Direct Submission
TITLE Submitted (28-OCT-1999) to the DDBJ/EMBL/GenBank databases.
JOURNAL Masahira Hattoni, The Institute of Physical and Chemical Research
(RIKEN), Genomic Sciences Center (GSC), Kitasato Univ., 1-15-1
Kitasato, Sagamihara, Kanagawa 228-8555, Japan
(E-mail: hattoni@gscc.riken.go.jp, URL: http://hgp.gscc.riken.go.jp/)
Tel: 81-42-778-9923 Fax: 81-42-778-9924
COMMENT On Feb 18, 2000 this sequence version replaced gi:6172113.
----- Genomic Center
Center: RIKEN Genomic Sciences Center (GSC)
Center code: RIKEN

```


* consists of 9 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

1 6172: contig of 6172 bp in length
* gap of unknown length
* 6173 13947: contig of 7775 bp in length
* gap of unknown length
* 13948 21015: contig of 7068 bp in length
* gap of unknown length
* 21016 42198: contig of 21183 bp in length
* gap of unknown length
* 42199 65368: contig of 23170 bp in length
* gap of unknown length
* 65369 93602: contig of 28234 bp in length
* gap of unknown length
* 93603 124558: contig of 30956 bp in length
* gap of unknown length
* 124559 161247: contig of 36689 bp in length
* gap of unknown length
* 161248 200965: contig of 39718 bp in length.

FEATURES
Source

1-200965
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="3p"
/clone="RP11-1015N5"

BASE COUNT 58185 a 40692 c 41385 g 60699 t 4 othera
ORIGIN

Query Match 1.38; Score 37; DB 53; Length 200965;
Best Local Similarity 46.18; Pred. No. 75;
Matches 124; Conservative 0; Mismatches 145; Indels 0; Gaps 0;
OY 1427 caccgaagcttaccgggtcttcttattccttacttgagctgagagcttccagcagggccgg 1486
DB 20845 GCGCGAGCTTAGAGTCCGGCCCGGGCCCACTCCCTCACGGGCCCCCGGGCGCAGCGG 20786
OY 1487 tatcaaggtccattcatcagcagtgatcgagcagcggttctgtaacttgggaagctc 1546
DB 20785 CCGAA 20726
OY 1547 tcaacttccagattgatagggtccgcagagctggcgagctgctctgtttaccacaaa 1606
DB 20725 TAAACGGGGTGATGGTGGAGGGGTTGGCGCTCGGGTAGTGGCGGGCCCGGACTGGAGTG 20666
OY 1607 gtggcaggtgcatctgacgccccagagctctcggggtaccctccacacacacat 1666
DB 20665 GGGGAAGCAGGAGCG 20606
OY 1667 cgcgggggattagatgcagacttcgaca 1695
DB 20605 GGGGTGGTGGTGGGACCAAGGCA 20577

Search completed: May 17, 2000, 06:18:16
Job time: 3954 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 17, 2000, 04:40:51 ; Search time: 92.13 seconds
(without alignments)
7959-546 Million cell updates/aec

Title: US-09-151-189-1
Perfect score: 2931
Sequence: 1 cgcggacagccctggagaga.....aaatggcgatagatatcc 2931

Scoring table: IDENTITY_NUC
Gapop 10.0, Gapext 1.0
Searched: 311585 seqs, 125096042 residues
Total number of hits satisfying chosen parameters: 623170

Minimum DB seq length: 0
Maximum DB seq length: 1000000
Post-processing: Minimum Match 0%
Listing first 45 summaries

Database: N_Geneseq_36.4

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	64.6	2.2	1794	1 V56020	Haloperoxidase, CP
2	62.8	2.1	1791	1 V56021	Haloperoxidase, CP
3	35	1.2	5280	1 V74855	Staphylococcus aur
4	34.8	1.2	799	1 X07126	Staphylococcus aur
5	34.8	1.2	799	1 X07127	Staphylococcus aur
6	34.8	1.2	6022	1 V74403	Staphylococcus aur
7	34.2	1.2	1012	1 V72124	Human catalytic te
8	34.2	1.2	2848	1 V27872	Human telomerase p
9	34.2	1.2	3798	1 V27876	Human telomerase p
10	34.2	1.2	3855	1 V22382	Human telomerase r
11	34.2	1.2	3955	1 V22379	Human telomerase r
12	34.2	1.2	3972	1 V72125	Human catalytic te
13	34.2	1.2	4023	1 V50320	Human telomerase g
14	34.2	1.2	4037	1 V22428	Human telomerase r
15	34.2	1.2	4042	1 V72117	Human telomerase r
16	34.2	1.2	4335	1 V16979	Human catalytic te
17	34.2	1.2	50341	1 V22674	Human telomerase r
18	34.2	1.2	52297	1 T51411	DNA sequence of a
19	34.2	1.2	52298	1 Q47357	Mycobacteriophage
20	34	1.2	4762	1 T94052	L5 mycobacteriophage
21	34	1.2	5423	1 V74335	Human multilocus re
22	33.6	1.1	384	1 O60279	Staphylococcus aur
23	33	1.1	9170	1 O02829	Human brain expres
24	32.8	1.1	5137	1 V74528	DNA complementary
25	32	1.1	662	1 V52558	Staphylococcus aur
26	32	1.1	704	1 V10553	Human native hepat
27	32	1.1	704	1 V10553	Stealth virus nucl
28	31.8	1.1	2055	1 V11376	Mycobacterium spec
29	31.8	1.1	2061	1 X34159	Mycobacterium spec
30	31.8	1.1	7183	1 Q06074	Sequence encoding
31	31.8	1.1	7184	1 Q07050	ptx gene of Pasteu
32	31.2	1.1	1998	1 Q11127	Sequence encoding
33	31.2	1.1	2074	1 T39790	Mouse SH3P8 gene.
34	31	1.1	1150	1 T03650	31-O-desmethyl-PK5

C 35	31	1.1	3817	1	T72814	Human SPT5 CDNA.. N
36	31	1.1	49272	1	V35000	Mycobacteriophage
37	30.8	1.1	813	1	T13450	Lung cancer specif
38	30.8	1.1	822	1	T13451	Lung cancer specif
39	30.8	1.1	986	1	T13449	Lung cancer specif
40	30.8	1.1	986	1	T13454	HCAVIII phosphoryl
41	30.8	1.1	1104	1	T15448	Lung cancer specif
42	30.8	1.1	2134	1	T15452	Lung cancer specif
43	30.8	1.1	9515	1	Q55145	Pseudomonas aerugin
44	30.6	1.0	1600	1	Q85770	Low affinity nerve
C 45	30.4	1.0	829	1	X00723	Human secreted pro

ALIGNMENTS

RESULT 1

V56020
ID V56020 standard; CDNA: 1794 BP.
AC V56020;
DT 08-DEC-1998 (first entry)
KW Haloperoxidase, CP:BP01 encoding CDNA.
KW Haloperoxidase, enzyme: CP:BP01; CP:BP02; vanadium; halide specificity:
KW Bromine; iodine; ds.
OS Corallina pillulifera.
PR Key Location/Qualifiers
FT CDS 1..1794
FT /*tag= a
FT /product= "Haloperoxidase, CP:BP01"
FT /note= "the stop codon is not indicated"
J10248591-A.
22-SEP-1998.
PD 06-MAR-1997; J70539.
PF 06-MAR-1997; J70539.
PR (SAKA) OTSUKA PHARM CO LTD.
PA WPI: 98-560733/48.
DR P-PSDB; W80550.
DK New haloperoxidase gene - and corresponding vector, transformed host cell and method of preparation
FT Claim 2: Pages 13-15; 20pp: Japanese.
CC This CDNA encodes a haloperoxidase enzyme. The haloperoxidase genes encoding the enzymes CP:BP01 and CP:BP02 are isolated from Corallina pillulifera, a heat cell transformed with a vector containing the haloperoxidase genes can be used for the recombinant production of the enzyme. The haloperoxidase of the invention are vanadium-dependent and have halide specificity of bromine and iodine.
CC Sequence 1794 BP; 395 A; 481 C; 505 G; 413 T;

Query Match 2.2a; Score 64.6; DB 1; Length 1794;
Best Local Similarity 49.2a; Pred. No. 2a-10;
Matches 203; Conservative 0; Mismatches 204; Indels 6; Gaps 1;

QY	1736	ggagataaatggggcgagatcccaacacgaggtcacctaccctcttccacaagctat	1795
DB	1359	GCAGAACATAGTCACGAGATCTCTGCTTCATCTCTGTCGCGATGGCATT	1418
QY	1796	cccaatggatgcgcacgcacccctccctcccgccgcacccacccaataatgagc	1855
DB	1419	CGCGGAGGCGAGCCATTCATCTGTCGACGAGCGCCACGCTGGTGTGTCGCGC	1478
QY	1856	atttgccacagttctgaagccctcattgttcctagatggggaggtgagtgctccctaa	1915
DB	1479	ANGTGTGACGATCTGAGGCGGTCTTCGACTCC-----GSCATCGAGATCGATCAGGT	1532
QY	1916	ccccgtgtccacagcgtacgcgcctggagcctgaactcaactcgaagggcgatgcctac	1975
DB	1535	GTTCCAGGTCACAAAGATGAGGACAGCTTGTGAAGTCGCTTTCAGGAGACTCTCAC	1592
QY	1976	atatgagggagagatcacacagctcggtcacctcatttgaggagcagatgctgg	2035
DB	1593	TGTTCCCGTGAAATTGAACAAGCTCGCCGACATATTCGGATCGCGGTACATGCGCAGG	1652

QY	2036	catcactatcgggttcgagcggatataccaagggctacttctctcgagagacaaactctgacy	2095
DB	1653	TGTTCACTACTCTCTGACCAAGTTCGATGCTCACTTCGCTCGTGGAGCAGGTCGGATTGG	1712
QY	2096	aacacttcaccaggagctgatgacgttcgcgcgaggaagccaccttgaattcc	2148
DB	1713	AATCTGGACAGCAAAAGTCGTGACGTATGGCGAGAACTCTCTCTCAACTGC	1765
RESULT 2			
V56021			
ID	V56021	standard; cDNA: 1791 BP.	
AC	V56021		
AD	09-DEC-1998	(first entry)	
DE	Haloperoxidase, CP.BP02 encoding cDNA.		
DE	Haloperoxidase, CP.BP02 encoding cDNA.		
KK	bromine; iodine; ds;		
KK	Corallina pilulifera.		
OS	Corallina pilulifera.		
OS	Key	Location/Qualifiers	
FT	CDS	1..1791	
FT		/*tag..a	
FT		/product= "Haloperoxidase, CP.BP02"	
FT		/note= "the stop codon is not indicated"	
FT			
FT	J10248581-A.		
FT	22-SEP-1998		
PP	06-MAR-1997;	070539.	
PP	06-MAR-1997;	JP-070539.	
PA	(SAKA.) OTSUKA PHARM CO LTD.		
PA	WPI: 98-560733/48.		
P	P-PSDB: X80551.		
PR	New haloperoxidase gene - and corresponding vector. transformed host		
PR	cell and method of preparation		
PR	Citium 2, pages 15-16; JOP; Japanese.		
BS	This cDNA encodes a haloperoxidase enzyme. The haloperoxidase genes		
BS	encoding the enzymes CP.BP01 and CP.BP02 are isolated from Corallina		
CCC	pilulifera. A host cell transformed with a vector containing the		
CCC	haloperoxidase genes can be used for the recombinant production of the		
CCC	enzyme. The haloperoxidases of the invention are vanadium-dependent and		
CCC	have halide specificity of bromine and iodine.		
SC	Sequence 1791 BP: 390 A: 459 C: 507 G: 435 T:		
Query Match 2.18; Score 62.8; DB 1: Length 1791;			
Best Local Similarity 46.78; Pred. NO. 7.7e-10;			
Matches 239; Conservative 0; Mismatches 267; Indels 6; Gaps			
QY	1666	tcgcgggggattcagatgcagacttcgacatccctcttcgaaaatgagactcttga	1725
DB	1286	TCCCTGAAGTTAGTGAACCTGGTCGAGGAGCTCTCATCAATCTCGATGACGTTGCTGAGA	1345
QY	1726	aagctgtgcgagagataaatggcgcgagatcccaacaacgaaggtcacctacctcttc	1785
DB	1346	GCAATGAAACAGAACAGAGGGCTGACGGGATTGTGAGCCGGATAAATCATTTCTGTCG	1405
QY	1786	cacaagctatccaaatgggatcgcaacgcacacctcttcacgcgcgcagctaccc	1845
DB	1406	CGATGGCAATTCCGAGGGGCGCCCATTCATCCGTCTATGGAAGTGGCCAGCTGTGG	1455
QY	1846	aaaatgagacatttcgcaacatttgagggccctcatctggctagatcgggggaggtgaat	1905
DB	1456	TTGCTGGCGCATGTGTGACAAATCTCANGCGGTCTTCAGCCCA-----CTTCCAGA	1519
QY	1906	gtctcccttaaccccgctgttcccaagcgtatgacggctcggaactaatcaattcgaaaggg	1965
DB	1520	TGATACAGGTGTCGAGGTGCACTACTGATGACGAAGCTGTGAAGTCGCTCTTCAAGG	1579
QY	1966	catgccttacatgatggggagagatcaacagctgcgcggtcaacgctgcatttgggagagc	2025
DB	1580	GAATCTCTACTGTTCGGGTGAATGTAAACAAAGTCGCGAGAAATGTGCATCGGGCGGA	1639
QY	2026	agatgctggggcatccactatcggttcgacgggtatccaaagcctactcttcgagagacaa	2085
DB	1640	ACATGGCGGGTGTTCATCTACTCTCTGACAGTGTGAGTCACTGCTACTCGGTGAACAGA	1699

QY	2086	toactgtacgaacacttcaccagagagctgtatgactgtcccgagagagccactttgaat	2145
DB	1700	TTGCGATTGAGTCATCTGGAGGAGCAGAGCTGACGTGATCGGAGAACCTCTCTCAATT	1759
QY	2146	tccgtctattaccgcgagaggtcatcaaacct	2177
DB	1760	TGCCGAAATTCGACGGAACCTACAATCCAGATT	1791
RESULT	3		
CV	V74855/C		
ID	V74855	standard; DNA: 5280 BP.	
AC	V74855;		
DE	16-MAR-1999	(first entry)	
DT	Staphylococcus aureus	contig SEQ ID #544.	
DE	Computer readable medium; vaccine; S.aureus infection; immunodetection;		
KW	cellulitis; eye lid infection; food poisoning; osteomyelitis; therapy;		
KW	skin infection; surgical wound infection; scalded skin syndrome;		
KW	toxic shock syndrome; ds.		
OS	Staphylococcus aureus		
FT	Key	Location/Qualifiers	
FT	misc_feature	541..600	
FT	FT	/*tag= a	
FT	FT	/note=	"these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"
FT	FT	2341..2400	
FT	FT	/*tag= b	
FT	FT	/note=	"these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"
FT	FT	4141..4200	
FT	FT	/*tag= c	
FT	FT	/note=	"these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"
EP	EP-786519-A2.		
PN	30-JUL-1997.	100117	
PD	07-JAN-1997.	US-009861	
PR	05-JAN-1996.	US-009861	
PP	(HOMA-) HUMAN GENOME SCI INC.		
PI	Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA, Rosen CA;		
PI	WPI; 97-374922/35.		
DR	Polynucleotide(s) and proteins derived from Staphylococcus aureus		
PT	stored on computer readable medium and used in the production of		
PT	anti-S.aureus vaccines		
PT	Claim 1; Page 1479-1482; 3271pp; English.		
PS	This sequence represents one of 5191 Staphylococcus aureus DNA sequences		
CC	of the invention. The DNA sequences are recorded on a computer readable		
CC	medium, preferably selected from a floppy or hard disk, random access		
CC	memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using		
CC	the S.aureus DNA sequences, allow putative functions to be assigned so		
CC	that protein-encoding or regulatory regions of commercial, therapeutic or		
CC	industrial importance can be obtained. Specifically, sequences which are		
CC	likely to encode antigens have been identified and these polypeptides can		
CC	be used in a vaccine composition against S.aureus infection. The		
CC	polypeptides can also be used in a kit for the immunodetection of		
CC	S.aureus in a sample. S.aureus is implicated in numerous human diseases,		
CC	including cellulitis, eye lid infections, food poisoning, osteomyelitis,		
CC	skin and surgical wound infections, scalded skin syndrome, toxic shock		
CC	syndrome, etc. Organisms transformed with the DNA sequences can be used		
CC	for recombinant production of the polypeptides. The new DNA sequences		
CC	(and their fragments) are useful as primers or probes for isolating		
CC	homologues of any of the S.aureus DNA sequences contained on the		
CC	computer readable medium.		
CC	sequence 5280 BP: 2057 A: 744 C: 854 G: 1442 T:		

Db 655 TATTGTCATGTTACACATGCTAGTAGTGGCTATATATAGGACCATCAGAGCATAC 712

RESULT 6

W74403

AC W74403 standard; DNA: 5022 BP.

DE 16-MAR-1999 (first entry)

DE Staphylococcus aureus confg SEQ ID #92.

DE Computer readable medium; vaccine; S.aureus infection; immunodetection;

KW cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;

KW skin infection; surgical wound infection; scalded skin syndrome;

KW toxic shock syndrome; ds.

OS Staphylococcus aureus.

PH Key

FT misc_feature

FT Location/Qualifiers

FT 421..480

FT /*tag= a

FT /*note= "these bases represent a line of missing text in

FT the sequence listing in the specification. They

FT are included to maintain the nucleotide numbering

FT given in the specification for this DNA sequence"

FT misc_feature 2221..2280

FT /*tag= b

FT /*note= "these bases represent a line of missing text in

FT the sequence listing in the specification. They

FT are included to maintain the nucleotide numbering

FT given in the specification for this DNA sequence"

FT misc_feature 4021..4080

FT /*tag= c

FT /*note= "these bases represent a line of missing text in

FT the sequence listing in the specification. They

FT are included to maintain the nucleotide numbering

FT given in the specification for this DNA sequence"

FT misc_feature 5821..5880

FT /*tag= d

FT /*note= "these bases represent a line of missing text in

FT the sequence listing in the specification. They

FT are included to maintain the nucleotide numbering

FT given in the specification for this DNA sequence"

EP-786519-A2.

30-JUL-1997. 100117.

07-JAN-1997. US-009861.

05-JAN-1996: US-009861.

(HUMAN) HUMAN GENOME SCI INC.

PI Resen GS; Choi GH, Dillon PV, Fannon MR, Kunsch CA,

PI WPI; 97-374922/35.

DR Polynucleotide(s) and proteins derived from Staphylococcus aureus

PT stored on computer readable medium and used in the production of

PT anti-S.aureus vaccines

PS Claim 1: page 569-573; 327pp: English.

CC This sequence represents one of 5191 Staphylococcus aureus DNA sequences

CC of the invention. The DNA sequences are recorded on a computer readable

CC medium, preferably selected from a floppy or hard disk, random access

CC memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using

CC the S.aureus DNA sequences allows putative functions to be assigned so

CC that protein-encoding or regulatory regions of commercial, therapeutic or

CC industrial importance can be obtained. Specifically, sequences which are

CC likely to encode antigens have been identified and these polypeptides can

CC be used in a vaccine composition against S.aureus infection. The

CC polypeptides can also be used in a kit for the immunodetection of

CC S.aureus in a sample. S.aureus is implicated in numerous human diseases,

CC including cellulitis, eyelid infections, food poisoning, osteomyelitis,

CC skin and surgical wound infections, scalded skin syndrome, toxic shock

CC syndrome, etc. Organisms transformed with the DNA sequences can be used

CC for recombinant production of the polypeptides. The new DNA sequences

CC (and their fragments) are useful as primers or probes for isolating

CC homologues of any of the S.aureus DNA sequences contained on the

CC computer readable medium.

CC Sequence 6022 BP: 2021 A: 993 C: 805 G: 1960 T;

Query Match 1.24; Score 34.8; DB 1; Length 6022;

Best Local Similarity 46.58; Pred No. 2.1;

Matches 111; Conservative 0; Mismatches 127; Indels 0; Gaps 0;

OY 340 gggagaaatcactaggttttcccaaggagagcccttataatcaagttacttgaaggt 399

DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

DB 604 GGCACAAATACACTATATATACAGAGCTAGTAATATACATCAGAAATGGAATGAT 663

OY 400 ctgagactctggagaaaggttaagcttcccaacggacgacatatacggtctccaagatctgg 459

DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

DB 664 TCACATTCAAAAGAGTATATACATTCATACACGCGCAAAATATTAACAATATATCTGC 723

OY 460 Gtaaaattatggccaggttcccatcgctaccgcttggctggctggctggctggctggctgg 519

DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

DB 724 ATTCAATATATATATTTGATGCGAGATGTTAAATATATATAGCATTTGGTAAATGACCAATGA 783

OY 520 gtttggcattcgacgaggtccacagccaggtgttttccctcgaggagacacacacacac 577

DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

DB 784 TATTGTCATGTTACACATGCTAGTAGTGGCTATATATAGGACCATCAGAGCATAC 841

RESULT 7

V72124

ID V72124 standard; cDNA: 1012 BP.

AC V72124;

DE 24-MAY-1999 (first entry)

DE Human catalytic telomerase sub-unit RACE-generated cDNA #3

KW Human; catalytic telomerase subunit; therapy; diagnosis; htc; assay;

KW modulator; treatment; inhibit; cellular disorder; death; defect; cancer;

KW ageing; antisense; neoplastic cell; telomerase-related condition; RACE;

KW tumour cell; ss.

OS Homo sapiens.

PD 30-DEC-1998.

PF 09-JUN-1998; E03468.

PR 14-APR-1998; DE-016496.

PR 20-JUN-1997; DE-026329.

PR 26-MAR-1998; DE-013274.

PA (FARB) BAYER AG.

PI Hagen G, Siegmund H, Weichel W, Wick M, Zubov D;

DR WPI; 99-081276/07.

PT New catalytically active subunit of human telomerase - used in the

PT modulation of telomerase activity, particularly for treating cancer

PT and ageing

PS Example 6; fig 10: 76pp: German.

CC This sequence represents a RACE-generated fragment of a novel human

CC catalytic telomerase sub-unit (hTC). The hTC protein can be used

CC in screening assays to identify modulators of telomerase and to treat or

CC inhibit cellular disorders, death, defects and/or other pathological

CC processes involving telomerase, particularly cancer and ageing (also

CC suitable for this are agents that stimulate, inhibit or mimic the

CC activity of the subunit). Antisense nucleic acids inhibit telomerase

CC action (by binding to specific mRNA), particularly in neoplastic cells

CC and may be expressed in vivo. Antibodies and fragments of the protein,

CC used as probes or primers, are used to diagnose telomerase-related

CC conditions (especially neoplasia) by (i) detecting abnormal levels of

CC the subunit protein in body fluids or tissues or (ii) by measuring the

CC amount of the encoding nucleic acid. Expression of the nucleic acid

CC encoding the subunit mRNA is confined to tumour cells, in contrast to

CC the ubiquitous expression of the telomerase RNA subunit.

CC Sequence 1012 BP: 159 A: 359 C: 320 G: 174 T;

Query Match 1.28; Score 34.2; DB 1; Length 1012;

Best Local Similarity 49.74; Pred No. 1;

Matches 87; Conservative 0; Mismatches 88; Indels 0; Gaps 0;

OY 1495 gtccattcaccagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcagtcact 1554

DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

DB 223 GACCAAGCATCT 282

OY 1555 teagattgatagtgccgc 1614

DB ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

DB 283 CAGCT 342

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: May 17, 2000, 04:39:38 ; Search time 56.46 Seconds
(without alignments)
6591.506 Million cell updates/sec

Title: US-09-151-189-1
Perfect score: 2931
Sequence: 1 cgcggacagccttggaga.....aaatggcgatagattccc 2931

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 226296 seqs, 63486255 residues

Total number of hits satisfying chosen parameters: 452592

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database : Issued_Patents_NA.*
1: /cgn2.6/prodata/1/ina/5A.COMB.seq.*
2: /cgn2.6/prodata/1/ina/5B.COMB.seq.*
3: /cgn2.6/prodata/1/ina/5C.COMB.seq.*
4: /cgn2.6/prodata/1/ina/5D.COMB.seq.*
5: /cgn2.6/prodata/1/ina/6.COMB.seq.*
6: /cgn2.6/prodata/1/ina/PCUS.COMB.seq.*
7: /cgn2.6/prodata/1/ina/backfiles.seq.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	35	1.2	7218	1	US-08-232-463-14
2	34.2	1.2	50341	2	US-08-247-901C-1
3	34.2	1.2	50341	4	US-09-075-904-1
C 4	32	1.1	704	2	US-08-463-115-29
C 5	32	1.1	704	2	US-08-463-388-29
C 6	31.2	1.1	1998	7	5212296-8
C 7	31	1.1	1150	1	US-08-264-861A-10
8	31	1.1	1150	6	PCT-US95-07784-10
9	31	1.1	2156	4	US-08-959-011-2
10	31	1.1	49272	2	US-08-614-770A-1
11	30.8	1.1	813	1	US-08-276-919-9
12	30.8	1.1	813	2	US-08-776-088-12
13	30.8	1.1	813	6	PCT-US95-09145A-12
14	30.8	1.1	822	2	US-08-776-088-14
15	30.8	1.1	822	6	PCT-US95-09145A-14
16	30.8	1.1	986	1	US-08-276-919-3
17	30.8	1.1	986	1	US-08-276-919-12
18	30.8	1.1	986	2	US-08-776-088-3
19	30.8	1.1	986	2	US-08-776-088-17
20	30.8	1.1	986	6	PCT-US95-09145A-3
21	30.8	1.1	986	6	PCT-US95-09145A-17
22	30.8	1.1	1104	1	US-08-276-919-1
23	30.8	1.1	1104	2	US-08-776-088-1
24	30.8	1.1	1104	6	PCT-US95-09145A-1
25	30.8	1.1	1438	4	US-08-845-742-1
26	30.8	1.1	2134	2	US-08-776-088-5
27	30.8	1.1	2134	6	PCT-US95-09145A-5

28	30.8	1.1	9515	2	US-08-920-812-13	Sequence 13, Appl
29	30.8	1.1	9515	2	US-08-920-827-13	Sequence 13, Appl
30	30.8	1.1	9515	4	US-08-921-177-13	Sequence 13, Appl
31	30.8	1.1	9515	3	US-08-920-828-13	Sequence 13, Appl
32	30.8	1.1	9515	3	US-08-920-828-13	Sequence 13, Appl
C 33	30.2	1.0	336	4	US-07-814-220-3	Sequence 3, Appl
C 34	30.2	1.0	336	4	US-07-814-220-4	Sequence 4, Appl
C 35	30.2	1.0	336	4	US-07-812-421-3	Sequence 3, Appl
C 36	30.2	1.0	336	4	US-07-812-421-4	Sequence 4, Appl
37	30.2	1.0	1290	5	US-08-745-977-3	Sequence 3, Appl
38	30.2	1.0	1290	5	US-09-040-699A-3	Sequence 3, Appl
39	30	1.0	434	5	US-08-866-340-25	Sequence 25, Appl
C 40	30	1.0	1635	1	US-08-133-347-1	Sequence 1, Appl
C 41	30	1.0	1635	1	US-08-133-347-3	Sequence 3, Appl
C 42	30	1.0	7218	1	US-08-232-463-14	Sequence 14, Appl
C 43	29.6	1.0	1662	2	US-08-651-573-1	Sequence 1, Appl
44	29.6	1.0	1662	5	US-09-066-543-1	Sequence 1, Appl
45	29.6	1.0	2277	2	US-08-676-967-2	Sequence 2, Appl

ALIGNMENTS

RESULT 1
US-08-232-463-14
Sequence 14, Application US/08232463
Patent No. 5670367
GENERAL INFORMATION:
APPLICANT: DORNER, F.
APPLICANT: SCHEIFLINGER, F.
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patentia Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 IMM
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: PT29pt-F15
US-08-232-463-14

Query Match 1.28; Score 36; DB 1; Length 7218;
 Best Local Similarity 2.48; Pred No. 0.58; Mismatches 163; Yndels 0; Gaps 0;
 Matches 9; Conservative 208;
 QY 868 cggagagacccggcgctgcctcgaatccatccatcgctggttgccatcacatccg 927
 Db 1051 CGAGGAGCTTGCAGATTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTTT 1110
 QY 928 gtccgcatctcggcagcaatacccccgggtacacccctctctcctcagctcg 987
 Db 1111 YY 1170
 QY 988 ccgctcagttggcggagctactgagcgctgcccagggatgaccccttatgag 1047
 Db 1171 YY 1230
 QY 1048 atggcagcagcaaatatacactaccgagcagcgaacccctcgtggaatggag 1107
 Db 1231 YY 1290
 QY 1108 caaatcggagcgcgtgctgaggggtccggtgagcggttccagctct 1167
 Db 1291 YY 1350
 QY 1168 tcgagcagcctggtgtgtaaacggggcccttctcctcagctcgtcgaca 1227
 Db 1351 YY 1410
 QY 1228 gcttcacacacgcgctatt 1247
 Db 1411 YYYYYYYYYYYYYYYYYYYY 1430

RESULT 2
 US-08-247-901C-1
 ; Sequence 1, Application US/08247901C
 ; Patent No. 5750384
 ; GENERAL INFORMATION:
 ; APPLICANT: Jacobs et al
 ; TITLE OF INVENTION: L5 SHUTTLE PHASMIDS
 ; NUMBER OF SEQUENCES: 1
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Amster, Rothstein & Ebenstein
 ; STREET: 90 Park Avenue
 ; CITY: New York
 ; STATE: New York
 ; COUNTRY: U.S.A.
 ; ZIP: 10016
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: 3.5 inch 1.44 Mb storage diskette
 ; OPERATING SYSTEM: MS-DOS
 ; SOFTWARE: Word Processor (ASCII)
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/247.901C
 ; FILING DATE: May 23, 1994
 ; CLASSIFICATION: 435
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/057.531
 ; FILING DATE: April 29, 1993
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Bogosian, Elizabeth A
 ; REGISTRATION NUMBER: 39,911
 ; REFERENCE/DOCKET NUMBER: 96700/273
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (212) 697-5995
 ; TELEFAX: (212) 286-0854 or 286-0082
 ; TELETYPE: TWX 710-581-4766
 ; INFORMATION FOR SEQ ID NO: 1:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 50341
 ; TYPE: nucleic acid
 ; STRANDEDNESS: single

TOPOLOGY: linear
 MOLECULE TYPE: L5 shuttle phasmid sequence
 DESCRIPTION: No
 HYPOTHETICAL: No
 ANTI-SENSE:
 FRAGMENT TYPE:
 ORIGINAL SOURCE:
 ORGANISM: L5 mycobacteriophage
 STRAIN:
 INDIVIDUAL ISOLATE:
 DEVELOPMENTAL STAGE:
 HAPLOTYPE:
 TISSUE TYPE:
 CELL LINE:
 ORGANELLE:
 IMMEDIATE SOURCE:
 POSITION IN GENOME:
 CHROMOSOME/SEGMENT:
 FEATURE:
 NAME/KEY:
 LOCATION:
 IDENTIFICATION METHOD:
 OTHER INFORMATION:
 PUBLICATION INFORMATION: No. 5750384e
 AUTHORS:
 TITLE:
 JOURNAL:
 VOLUME:
 PAGES:
 DATE:
 DOCUMENT NUMBER:
 FILING DATE:
 PUBLICATION DATE:
 RELEVANT RESIDUES IN SEQ ID NO:
 US-08-247-901C-1
 Query Match 1.28; Score 34.2; DB 2; Length 50341;
 Best Local Similarity 50.94; Pred. No. 8.8;
 Matches 81; Conservative 0; Mismatches 78; Indels 0; Gaps 0;
 QY 1296 gtccgatttgcaggaatggtgcaattcagatggtgagccccggcgcccgagag 1355
 Db 8714 GTCGACTCAGTCGCCGAGCGCCAGCGGCTTCGCGCTCGCGGATCGCGAGAG 8773
 QY 1356 ttgacgagagctgctgttttttccgttaacgcccgcgacctggccaggtctctctg 1415
 Db 8774 GCTGACGAGAGCTGTGCGAGTGTGCGAGCGCCACCAACACCTCGACATCGAGGCCACTG 8833
 QY 1416 gacaatatcaacacccgagcttatcggggtctcttate 1454
 Db 8834 GCTACACCGAGCTTACGTTACGCGCGCTGTTTCATC 8872
 RESULT 3
 US-09-075-904-1
 ; Sequence 1, Application US/09075904
 ; Patent No. 5994137
 ; GENERAL INFORMATION:
 ; APPLICANT: Jacobs, et al.
 ; TITLE OF INVENTION: L5 SHUTTLE PHASMIDS
 ; NUMBER OF SEQUENCES: 1
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Amster, Rothstein & Ebenstein
 ; STREET: 90 Park Avenue
 ; CITY: New York
 ; STATE: New York
 ; COUNTRY: U.S.A.
 ; ZIP: 10016
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: 3.5 inch 1.44 Mb storage diskette
 ; STRANDEDNESS: single

OPERATING SYSTEM: MS-DOS
 SOFTWARE: Word Processor (ASCII)
 CURRENT APPLICATION DATA: 05/09/075.904
 APPLICATION NUMBER: 05/09/075.904
 FILING DATE: May 11, 1998
 CLASSIFICATION:
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: 08/247.901
 FILING DATE: May 23, 1994
 ATTORNEY/AGENT INFORMATION:
 NAME: Bogosian, Elizabeth A
 REGISTRATION NUMBER: 39,911
 REFERENCE/DOCKET NUMBER: 96700/475
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (212) 697-5995
 TELEFAX: (212) 286-0854 or 286-0082
 TELEX: TWX 710-581-4766
 INFORMATION FOR SEQ ID NO: 1:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 50341
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 MOLECULE TYPE: L5 shuttle phasmid sequence
 DESCRIPTION: L5 shuttle phasmid sequence
 HYPOTHEICAL: NO
 ANTI-SENSE:
 FRAGMENT TYPE:
 ORGANISM: L5 mycobacteriophage
 STRAIN:
 INDIVIDUAL ISOLATE:
 DEVELOPMENTAL STAGE:
 HAPLOTYPE:
 TISSUE TYPE:
 CELL TYPE:
 CELL LINE:
 ORGANELLE:
 IMMEDIATE SOURCE:
 POSITION IN GENOME:
 CHROMOSOME/SEGMENT:
 FEATURE:
 NAME/KEY:
 LOCATION:
 IDENTIFICATION METHOD:
 OTHER INFORMATION:
 PUBLICATION INFORMATION: No. 5994137e
 AUTHORS:
 TITLE:
 JOURNAL:
 VOLUME:
 PAGES:
 DATE:
 DOCUMENT NUMBER:
 FILING DATE:
 PUBLICATION DATE:
 RELEVANT RESIDUES IN SEQ ID NO:

US-09-075-904-1

Query Match 1.28; Score 34.2; DB 4; Length 50341;
 Best Local Similarity 50.9%; Pred. No. 8.8;
 Matches 81; Conservative 0; Mismatches 78; Indels 0; Gaps 0.
 Oy 1296 gtgatttgcagaaatggtgaacattcaatggtggaccocccgcccgcgcgaagag 1355
 Ob 8714 GTCGACTGAGTCGCGAGCGCCAGCGCGCTGCGAGGTTTCGCGCTCGCGATGCGGACG 8773
 Oy 1356 ttgacgagagcgtgcttttattcgttaacgcccgcgacctggccagggctctctctg 1415
 Db 8774 GCTGACGAGAGCTGTGCGAGTGTGCGAGGCAACAACTGACATCGAGGCAACCACTG 8833
 Oy 1416 gacaaatacaacacgagcttatcgcggtctcttctc 1454

Db 8834 GGTACACCGAGCGTTAGCTTACGCGCGCTCGTTCATC 8872
 RESULT 4
 US-08-463-115-29/c
 Sequence 29, Application US/08463115
 Patent No. 5703221
 GENERAL INFORMATION:
 APPLICANT: WILLIAM JOHN MARTIN
 TITLE OF INVENTION: ISOLATED STEALTH VIRUSES
 TITLE OF INVENTION: AND RELATED VACCINES
 NUMBER OF SEQUENCES: 104
 CORRESPONDENCE ADDRESS:
 ADDRESSEE: LYON & LYON
 STREET: 633 West Fifth Street
 CITY: Los Angeles
 STATE: California
 COUNTRY: U.S.A.
 ZIP: 90071-2066
 COMPUTER READABLE FORM:
 MEDIA TYPE: 3.5" Diskette, 1.44 Mb
 MEDIA TYPE: storage
 COMPUTER: IBM Compatible
 OPERATING SYSTEM: IBM P.C. DOS 5.0
 SOFTWARE: FastSeq Version 1.5
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/463.115
 FILING DATE: June 5, 1995
 CLASSIFICATION: 435
 PRIOR APPLICATION DATA:
 PRIOR APPLICATION DATA: including application
 PRIOR APPLICATION DATA: described below:
 APPLICATION NUMBER: 08/157.814
 FILING DATE: No. 5703221ember 23, 1993
 APPLICATION NUMBER: 07/887.502
 FILING DATE: May 22, 1992
 APPLICATION NUMBER: 07/704.814
 FILING DATE: May 23, 1993
 APPLICATION NUMBER: 07/763.039
 FILING DATE: September 20, 1991
 ATTORNEY/AGENT INFORMATION:
 NAME: Waizburg, Richard J.
 REGISTRATION NUMBER: 32,327
 REFERENCE/DOCKET NUMBER: 213/301
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (213) 489-1600
 TELEFAX: (213) 955-0440
 TELEX: 67-3510
 INFORMATION FOR SEQ ID NO: 29:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 704 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: single
 TOPOLOGY: linear
 FEATURE:
 OTHER INFORMATION:
 US-08-463-115-29

Query Match 1.18; Score 32; OB 2; Length 704;
 Best Local Similarity 57.7%; Pred. No. 2.2;
 Matches 56; Conservative 0; Mismatches 41; Indels 0; Gaps 0.
 Oy 1757 tcccaacaagaggtcacctacattcttccacaagctatccaagtggtatcgccaacga 1816
 Ob 215 TCCGCCACGAGTGTGACGGCTGCTCCCATAGGGTTGCATTGGGCTTCCCATGCC 156
 Oy 1817 cctctctaccctgcccggccagctaccacaatgga 1853
 Ob 155 GCTTTTACCCAGCGGCCGCGCTCTCGATTGCA 119

TELEPHONE: (908) 594-3904
TELEFAX: (908) 594-4720
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 1150 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
US-08-264-861A-10

Query Match 1.1%; Score 31; DB 1; Length 1150;
Best Local Similarity 52.8%; Pred. No. 6.7;
Matches 67; Conservative 0; Mismatches 60; Indels 0; Gaps 0;
QY 1153 cttctccagctctccagagcacttctgtgtttgaacagggccctttgtctctc 1212
DB 1007 CGATCTCCCTGTTCTAGCGCCGACGCGCGCCGACCGCGCGCTCCGACTCG 1066
QY 1213 agctctgtgaacagcttaccatgacgctattacggtcgaacagcagagacat 1272
DB 1067 TCAGTGACCGCTCCGCTCCCGCCGCGCTCCCGCGCGCTCCGAGGGCT 1126
QY 1273 tgcgccc 1279
DB 1127 CCGCGCC 1133

RESULT 8
PCT-US95-07784-10
Sequence 10, Application PC/TUS9507784
GENERAL INFORMATION:
APPLICANT: MOTAMED, HAIDEN
APPLICANT: SHAFIEE, ALI
TITLE OF INVENTION: EXPRESSION CASSETTES USEFUL IN
TITLE OF INVENTION: CONSTRUCTION OF INTEGRATIVE AND REPLICATIVE EXPRESSION
NUMBER OF SEQUENCES: 11
CORRESPONDENCE ADDRESS:
ADDRESSEE: J. ERIC THIES
STREET: P.O. BOX 2000, 136 E. LINCOLN AVENUE
CITY: RAHWAY
STATE: NJ
COUNTRY: USA
ZIP: 07065
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/07784
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: THIES, J. ERIC
REGISTRATION NUMBER: 35,382
REFERENCE/DOCKET NUMBER: 19132
TELECOMMUNICATION INFORMATION:
TELEPHONE: (908) 594-3904
TELEFAX: (908) 594-4720
INFORMATION FOR SEQ ID NO: 10:
SEQUENCE CHARACTERISTICS:
LENGTH: 1150 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: unknown
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO

PCT-US95-07784-10

Query Match 1.1%; Score 31; DB 6; Length 1150;
Best Local Similarity 52.8%; Pred. No. 6.7;
Matches 67; Conservative 0; Mismatches 60; Indels 0; Gaps 0;
QY 1153 cttctccagctctccagagcacttctgtgtttgaacagggccctttgtctctc 1212
DB 1007 CGATCTCCCTGTTCTAGCGCCGACGCGCGCCGACCGCGCGCTCCGACTCG 1066
QY 1213 agctctgtgaacagcttaccatgacgctattacggtcgaacagcagagacat 1272
DB 1067 TCAGTGACCGCTCCGCTCCCGCCGCGCTCCCGCGCGCTCCGAGGGCT 1126
QY 1273 tgcgccc 1279
DB 1127 CCGCGCC 1133

RESULT 9
US-08-959-011-2/c
Sequence 2, Application US/08959011
Patent No. 5932444
GENERAL INFORMATION:
APPLICANT: Hillman, Jennifer L.
APPLICANT: Shah, Purvi
APPLICANT: Corley, Neil C.
TITLE OF INVENTION: VACUOLAR ATPASE SUBUNIT AC45
NUMBER OF SEQUENCES: 3
CORRESPONDENCE ADDRESS:
ADDRESSEE: Incyte Pharmaceuticals, Inc.
STREET: 3174 Porter Drive
CITY: Palo Alto
STATE: CA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM compatible
OPERATING SYSTEM: DOS
SOFTWARE: FastSeq for Windows version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/959,011
FILING DATE: Herewith
CLASSIFICATION: 536
PRIOR APPLICATION NUMBER:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Billings, Lucy J.
REGISTRATION NUMBER: 36,749
REFERENCE/DOCKET NUMBER: PF-0412 US
TELECOMMUNICATION INFORMATION:
TELEPHONE: 650-855-0555
TELEFAX: 650-845-4166
TELEX:
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 2156 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
LIBRARY: CRBLNTOI
CLONE: 676592
US-08-959-011-2

Query Match 1.1%; Score 31; DB 4; Length 2156;
Best Local Similarity 46.9%; Pred. No. 10;
Matches 97; Conservative 0; Mismatches 110; Indels 0; Gaps 0;


```

Sequence 12, Application US/08776088
Patent No. 577379
GENERAL INFORMATION:
APPLICANT: Torczynski, Richard M.
APPLICANT: Bollon, Arthur P.
TITLE OF INVENTION: Lung Cancer Marker
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESSEE: SIDLEY & AUSTIN
STREET: 1201 Elm Street, Suite 4500
CITY: Dallas
STATE: TX
COUNTRY: US
ZIP: 75270-2197
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentia Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/776,088
FILING DATE: 19-JUL-95
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Eugenia S. Hansen
REGISTRATION NUMBER: 31,966
REFERENCE/DOCKET NUMBER: 10365/05011
TELECOMMUNICATION INFORMATION:
TELEPHONE: 214-981-3300
TELEFAX: 214-981-3400
INFORMATION FOR SEO ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 813 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: linear
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..813
US-08-776-088-12

Query Match 1.18; Score 30.8; DB 2; I
Best Local Similarity 49.4%; Pred. No. 6.1;
Matches 80; Conservative 0; Mismatches 82;

PY 1331 tgaaccccgccggccgcgaagaggttccttcgtggacaatatcaacaccggaag
DB 306 TGAGCACACCGTCGACGCGACAGCACTTCGCGCGCGAGCTGCACAT
PY 1391 cgacctggcccgaggttccttcgtggacaatatcaacaccggaag
DB 366 AGACCTTTATCTCGACGCGACACTGCACACACAGTCAGCAAG
PY 1451 tatcctacttgagctggagagccttcagcagcgcccggtatcaaa 14
DB 426 TGTTCTCATTTAGATGGGCTCCTTCATTCGTCCTATGACAA 14

RESULT 13.
CT-US95-09145A-12
Sequence 12, Application PC/TUS9509145A
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: Lung Cancer Marker
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESSEE: RICHARDS, MEDLOGK & ANDREWS
STREET: 1201 Elm Street, Suite 4500
CITY: Dallas
STATE: TX
COUNTRY: US

```

NAME: Eugenia S. Hansen
REGISTRATION NUMBER: 31,966
REFERENCE/DOCKET NUMBER: 10365/05011
TELECOMMUNICATION INFORMATION:
TELEPHONE: 214-981-3300
TELEFAX: 214-981-3400
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 822 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: linear
MOLECULE TYPE: CDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..822
US-08-776-088-14

Query Match 1.1%; Score 30.8; DB 2; Length 822;
Best Local Similarity 49.4%; Pred. No. 6.2;
Matches 80; Conservative 0; Mismatches 82; Indels 0; Gaps 0;
Qy 1331 tggaccccgccgcccgaaggttagacaagagctgcgtttttatccgtaacgccc 1390
Db 306 TGAGCACACCGTCAGCGGACGACGCTGCCCGGAGCTGCACATTGTCATTATAACTC 365
Qy 1391 cgaactggccagggtctcctctgtgacaatatcaacacggaagcttatcgcggtctct 1450
Db 366 AGACCTTTATCTGACGCCGACACTGCCAGCACTACACAGGCTGCTGCTGCTGCTG 425
Qy 1451 tatctacttgagctggagcccttcagcgcccggtatcaa 1492
Db 426 TGTCTCATGTGAGTGGGCTCTTCATCCGCTCTATGACAA 467

RESULT 15
PCT-US95-09145A-14
Sequence 14, Application PC/TUS9509145A
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION: Lung Cancer Marker
NUMBER OF SEQUENCES: 22
CORRESPONDENCE ADDRESS:
ADDRESSEE: RICHARDS, MEDLOCK & ANDREWS
STREET: 1201 Elm Street, Suite 4500
CITY: Dallas
STATE: TX
COUNTRY: US
ZIP: 75270-2197
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US95/09145A
FILING DATE:
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: John A. Harie
REGISTRATION NUMBER: 37,345
REFERENCE/DOCKET NUMBER: B35792CIPPCT
TELECOMMUNICATION INFORMATION:
TELEPHONE: 214-939-4500
TELEFAX: 214-939-4600
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 822 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: linear
MOLECULE TYPE: CDNA

FEATURE:
NAME/KEY: CDS
LOCATION: 1..822
PCT-US95-09145A-14

Query Match 1.1%; Score 30.8; DB 6; Length 822;
Best Local Similarity 49.4%; Pred. No. 6.2;
Matches 80; Conservative 0; Mismatches 82; Indels 0; Gaps 0;
Qy 1331 tggaccccgccgcccgaaggttagacaagagctgcgtttttatccgtaacgccc 1390
Db 306 TGAGCACACCGTCAGCGGACGACGCTGCCCGGAGCTGCACATTGTCATTATAACTC 365
Qy 1391 cgaactggccagggtctcctctgtgacaatatcaacacggaagcttatcgcggtctct 1450
Db 366 AGACCTTTATCTGACGCCGACACTGCCAGCACTACACAGGCTGCTGCTGCTGCTG 425
Qy 1451 tatctacttgagctggagcccttcagcgcccggtatcaa 1492
Db 426 TGTCTCATGTGAGTGGGCTCTTCATCCGCTCTATGACAA 467

Search completed: May 17, 2000, 05:51:46
Job time: 4328 sec

Thu May 18 22:21:28 2000

us-09-151-189-1.rst

Page 1

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sv model

Run on: May 17, 2000, 04:26:14 : Search time 944.91 Seconds
(without alignments)
12572.623 Million cell updates/sec.

Title: US-09-151-189-1
Perfect score: 2931
Sequence: 1 cgcggacaagccttggaaga.....aaatggcgatagatatcc 2931

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 4857316 seqs, 2026611650 residues

Total number of hits satisfying chosen parameters: 9714632

Minimum DB seq length: 0
Maximum DB seq length: 1000000

Post-processing: Minimum Match 0%
Listing first 45 summaries

Database :

EST:

1: em_est1:
2: em_est2:
3: em_est3:
4: em_est4:
5: em_est5:
6: em_est6:
7: em_est7:
8: em_est8:
9: em_est9:
10: em_est10:
11: em_est11:
12: em_est12:
13: em_est13:
14: em_est14:
15: em_est15:
16: em_est16:
17: em_est17:
18: em_est18:
19: em_est19:
20: gb_est1:
21: gb_est2:
22: gb_est3:
23: gb_est4:
24: gb_est5:
25: gb_est6:
26: gb_est7:
27: gb_est8:
28: gb_est9:
29: gb_est10:
30: gb_est11:
31: gb_est12:
32: gb_est13:
33: gb_est14:
34: gb_est15:
35: gb_est16:
36: gb_est17:
37: gb_est18:
38: gb_est19:
39: gb_est20:
40: gb_est21:
41: gb_est22:
42: gb_est23:
43: gb_est24:
44: gb_est25:

45: gb_est26:
46: gb_est27:
47: gb_est28:
48: gb_est29:
49: gb_est30:
50: gb_est31:
51: gb_est32:
52: em_est20:
53: em_est21:
54: em_est22:
55: em_est23:
56: em_est24:
57: em_est25:
58: em_est26:
59: gb_est33:
60: gb_est34:
61: gb_est35:
62: gb_est36:
63: gb_est37:
64: gb_est38:
65: em_est27:
66: em_est28:
67: em_est29:
68: em_est30:
69: gb_est39:
70: gb_est40:
71: gb_est41:
72: gb_est42:
73: gb_est43:
74: gb_est44:
75: em_est31:
76: em_est32:
77: em_est33:
78: em_est34:
79: gb_est45:
80: gb_est46:
81: gb_est47:
82: gb_est48:
83: gb_est49:
84: gb_est50:
85: gb_est51:
86: em_gss1:
87: em_gss2:
88: em_gss3:
89: em_gss4:
90: gb_gss5:
91: gb_gss6:
92: gb_gss7:
93: gb_gss8:
94: gb_gss9:
95: em_gss5:
96: em_gss6:
97: em_gss7:
98: em_gss8:
99: em_gss9:
100: em_gss10:
101: em_gss11:
102: gb_gss10:
103: gb_gss11:
104: em_gss12:
105: gb_gss12:
106: gb_gss13:
107: gb_gss14:
108: gb_gss15:
109: gb_gss16:

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result Query

No.	Score	Match	Length	DB	ID	Description
1	80.4	2.7	393	81	AW400475	Landisest
2	51.6	1.8	925	82	CNS0091P	Drosophila
3	40.4	1.4	942	82	CNS004NS	Drosophila
4	38.2	1.3	382	21	T62798	Yb9gill.r1
5	38	1.3	324	32	AA365809	EST76653
6	37.2	1.3	490	79	AW250569	2821781.5
7	37.2	1.3	504	28	AA062675	2801710.5
8	37.2	1.3	752	4	AT314531	u148b02.y
9	36.8	1.3	468	79	AT284526	L51.270.c
10	36.8	1.3	1044	83	CNS0150E	Drosophila
11	36.4	1.2	533	26	W45709	227403.81
12	36.4	1.2	610	108	AO576692	nubd0089L
13	36.2	1.2	327	36	C74220	C74220
14	36	1.2	334	34	AA491860	ne79902.s
15	36	1.2	446	80	AW232825	u074e09.y
16	36	1.2	557	37	AA171962	vt99ell.r
17	36	1.2	600	27	W84020	mr26908.r1
18	36	1.2	787	82	CNS010B7	Drosophila
19	35.8	1.2	488	43	AI205160	ao84d04.x
20	35.8	1.2	502	79	AW245215	2819614.5
21	35.6	1.2	446	25	D83844	R08HBC4877
22	35.6	1.2	1069	83	CNS015BE	Drosophila
23	35.4	1.2	1101	82	CNS00LJC	Drosophila
24	35.4	1.2	437	49	AI635496	ts65h04.x
25	35.4	1.2	645	82	CNS012T3	Drosophila
26	35.2	1.2	479	39	AA905715	o195el2.s
27	35.2	1.2	987	82	CNS0015S	Drosophila
28	35	1.2	469	70	AW148044	da03c09.x
29	35	1.2	1101	82	CNS00LXJ	Drosophila
30	34.8	1.2	247	80	AW407869	DI-RF-BL0
31	34.6	1.2	351	32	AA353352	EST61495
32	34.6	1.2	552	74	AO157443	nrb00009B
33	34.6	1.2	945	82	AV388992	Drosophila
34	34.6	1.2	945	82	CNS0092K	Drosophila
35	34.4	1.2	592	80	C99982	C99982
36	34.4	1.2	720	49	AI632354	tl22h09.x
37	34.4	1.2	775	49	AI632354	tl22h09.x
38	34.4	1.2	939	82	CNS00CKG	Drosophila
39	34.4	1.2	1101	82	CNS00CKG	Drosophila
40	34.2	1.2	343	24	N22510	yw39912.s1
41	34.2	1.2	384	60	AI800988	wg15b08.x
42	34.2	1.2	388	27	AA016068	ze32c08.x
43	34.2	1.2	428	44	AI311561	qo91c08.x
44	34.2	1.2	479	43	AI225585	u106b01.y
45	34.2	1.2	482	79	AW245544	2822724.5

ALIGNMENTS

RESULT 1
 LOCUS AW400475 393 bp mRNA EST 07-FEB-2000
 DEFINITION Landisest169est.L.digitata sporophyte Lambda ZapTI Laminaria
 digitata cDNA similar to vanadium bromoperoxidase, mRNA sequence.
 ACCESSION AW400475
 VERSION EST
 KEYWORDS
 SOURCE Laminaria digitata
 ORGANISM Laminaria digitata
 Eukaryota; stramenopiles; Phaeophyceae/Xanthophyceae group;
 Phaeophyceae; Laminariales; Laminariaceae; Laminaria.
 REFERENCE 1 (bases 1 to 393)
 Crepiniau, F., Roscoe, T., Kaas, R., Kloareg, B. and Boyen, C.
 Characterisation of complementary DNAs from the Expressed Sequence
 Tag analysis of life cycle stages of Laminaria digitata
 (Phaeophyceae)
 Unpublished
 JOURNAL On Dec 20, 1995 this sequence version replaced gi:1134111.
 CONTACT: Boyen C
 Centre d'Etudes Oceanologique et de Biologie Marine

CNRS-LPR9042, Universite P. & M. Curie
 BP74, F-29682 Roscoff cedex, France
 Tel: 33 2 98 29 23 32
 Fax: 33 2 98 29 23 24
 Email: est@b-roscoff.fr
 Location/Qualifiers
 1..393
 /organism="Laminaria digitata"
 /db_xref="taxon:80365"
 /clone_lib="L.digitata sporophyte Lambda ZapTI"
 /seq_stages="sporophyte"
 /lab_host="SOLR"
 Note="vector: paluescript SK+; Site_1: EcoRI; Site_2:
 XbaI; The Laminaria sporophyte library, constructed by T
 Roscoe and F Crepiniau, was oligo-(dT) primed and
 directionally cloned into a Uni-ZAPTM XR vector
 (Stratagene, La Jolla, CA, USA) using total mRNA from
 sporophytes harvested at 1 file de Sieck (F)."
 BASE COUNT 83 a 127 c 107 g 76 t
 ORIGIN

Query Match 2.7%; Score 80.4; DB 81; Length 393;
 Best Local Similarity 55.7%; Pred. No. 6e-13;
 Matches 215; Conservative 0; Mismatches 166; Indels 5; Gaps 3;
 QY 1035 ccccttatgagtagtcgacgacgaaattaccactaccgagcagccacacctcgctga 1094
 DB 13 CCCTTCCGGAGTTCGCAAGAACGACTACGTACGGCTCCGCCGACAGAACCTGCAGGCG 72
 QY 1095 atggggagctcccaatgtgacgacgctgctcgtatgggtccgatgctgagaccg 1154
 DB 73 CTGCGGGCGTTCAGGAGCTCAACATTCCTCCGGAGCGAG--CGAGGAAAGATCGACCT 129
 QY 1155 ttctccagctcttcgacgacacctctgtgtgttttaaaagggccctttctctcag 1214
 DB 130 GTCACTGATCTGTTCGACAGACCTTCGGCGGGTGTGACACCGCTCCCTCGTGTCTAG 189
 QY 1215 ctgctgtgacgcttaccactaccgactattacgacgacgacgacgacgacgacgac 1274
 DB 190 TCAATGCTGGGACATCTCTGATGACACTATCAAGTCAACCCGAGGCTGACCCCTT 248
 QY 1275 gcccccactgaactatattgtcgtatgttgaacgaatggctgaacattcagaatgttga 1334
 DB 249 ACCCCGCGCTGGGACTACATGACCGCTTCCAGCGCGTGGCTCGACGTACAGAACGG 307
 QY 1335 ccccccgc 1394
 DB 308 TTCCAAGCTTGAACCTTCTTCGACGAGGAGAACCCGCTTCATCCGCAACGCGGAGAC 367
 QY 1395 ctgaccaggtct 1420
 DB 368 CTGGCCACTATTCCTTTAGGNTAA 393

RESULT 2
 LOCUS CNS0091P 925 bp DNA GSS 03-JUN-1999
 DEFINITION Drosophila melanogaster genome survey sequence TE93 end of BAC #
 BACR19D16 of RPCI-98 library from drosophila melanogaster (fruit
 fly), genomic survey sequence.
 ACCESSION AL053013
 VERSION AL053013.1 GI:4934461
 KEYWORDS GSS
 SOURCE fruit fly
 ORGANISM Drosophila melanogaster
 Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
 Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
 Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
 REFERENCE 1 (bases 1 to 925)
 Genoscope.
 AUTHORS Direct Submission
 TITLE Direct Submission
 JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage

Seq primer: GGAACAGCTATGACCATG
Class: BAC ends
High quality sequence stop: 360.
Location/Qualifiers
1. 610

FEATURES

source

/organism="Oryza sativa"
/strain="Japonica"
/cultivar="Nipponbare"
/db_xref="taxon:4530"
/clone="fbab00891052"
/clone_lib="CGI Rice BAC Library"
/tissue_type="Leaf"
/lab_host="E. coli DR108"

/note="Vector: pBeloBAC11; Site_1: HindIII; Site_2: HindIII; Rice is one of two most popular grains in the world. Half of the world population especially those inhabiting highly populated areas of the humid tropics and subtropics, rely on rice as their primary source of carbohydrate. Monocotyledonous rice is a diploid plant (2n=24) with a haploid genome equivalent of 431 Mbp (Arumuganathan and Earle, 1991). The relatively small genome of rice, three times larger than that of Arabidopsis, makes it suitable for genomic studies. In order to facilitate positional cloning, physical mapping and genome sequencing of rice, we have constructed a BAC library from Oryza sativa, Nipponbare variety. The library contains 36,864 clones with an average insert size of 128.5 Kb providing 10.9 haploid genome equivalents. The deep coverage allows the isolation of a particular sequence with a probability of 99.9%. Two high density filters, each containing 18,432 clones (doubly spotted), represent the whole library for colony screening."

BASE COUNT
ORIGIN

146 a 176 c 173 g 114 t 1 others

Query Match 1.2%; Score 36.4; DB 108; Length 610;
Best Local Similarity 49.0%; Pred. No. 11;
Matches 97; Conservative 0; Mismatches 101; Indels 0; Gaps 0;

QY 855 gagagcgtgctgagagaccgagcgggctgctgctcgaatcctacccgctgctgttggcc 914
DB 336 GATGGCTTGCTAGCTGCTGCTCCCTGAGGGGAATGAACAGTAAGTCTGCTTTTCGCC 267
QY 915 atgacatataccggtccgctcattcggctcagcaataaccgctaccctccctcc 974
DB 266 ATATACCTTCGGCTTTAGCTCTCTCTGCTTGACCTTTTAACCTGACTCTTTCTTTCC 207
QY 975 tctcgtgagctccgctcagttggagagctatactggatggcgtggcagggatga 1034
DB 206 TTTCAGTGCTTACCTTGACCAACCCGACCCGATGATGATCGGCGACGTA 147
QY 1035 ccccttatgcagtatggc 1052
DB 146 TCGGCTGTTAGTACAGC 129

RESULT 13

C72420/c 327 bp mRNA EST 22-SEP-1997
DEFINITION C72420 Rice panicle at flowering stage Oryza sativa cDNA clone
E1585.1A, mRNA sequence.

ACCESSION C72420
VERSION C72420
KEYWORDS EST.

SOURCE

Oryza sativa

ORGANISM

Eukaryota; Viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; Liliopsida; Poales;
Poaceae; Oryza; (bases 1 to 327)

REFERENCE

AUTHORS

TITLE

Sasaki, T. and Yamamoto, K.
Rice cDNA from panicle at flowering stage

JOURNAL

COMMENT

Unpublished (1996)
On Sep 12, 1996 this sequence version replaced gi:1492260.
Contact: Takuji Sasaki
National Institute of Agrobiological Resources
Rice Genome Research Program
2-1-2 Kannondai, Tsukuba
Ibaraki,
Japan 305
Tel: 0298-38-7441
Fax: 0298-38-7468
Email: tsasaki@abrr.affrc.go.jp.
Location/Qualifiers
1. 327

FEATURES
source

/organism="Oryza sativa"
/strain="Nipponbare"
/db_xref="taxon:4530"
/clone="E1585.1A"
/clone_lib="Rice panicle at flowering stage"
/dev_stage="flowering stage"
/note="Organ: panicle; Rice cDNA from panicle at flowering stage"

BASE COUNT 47 a 113 c 111 g 53 t 3 others

ORIGIN

Query Match 1.2%; Score 36.2; DB 36; Length 327;
Best Local Similarity 52.7%; Pred. No. 8;
Matches 77; Conservative 0; Mismatches 69; Indels 0; Gaps 0;

QY 777 gatgaactggagcaggttaacgcagcagcagcttccttcctcagcagcagctcatttaac 836
DB 183 GAGCGCTTGAAGAGGAGAGCGGGGACGCGCTGCGCCCTGCTGCGACGCGCAGGAC 124
QY 837 ggaatttcagcatttgcgagcagctgctgcggagaccgagcggcctgcctgccttcacat 896
DB 123 GAGGTACGGCGCGCTGACCACTCGCACCGCTGCTGCTGCGACGCTGCGCACCTCGGT 64
QY 897 cctacgcgtgcgttggcattcgacat 922
DB 63 CGGCGCGCTGCGCGCGCTGCGGACGT 38

RESULT 14

AA491660

LOCUS

DEFINITION

AA491660

VERSION

AA491660.1

KEYWORDS

EST.

SOURCE

human

ORGANISM

Homo sapiens

REFERENCE

AUTHORS

TITLE

JOURNAL

COMMENT

On Sep 12, 1996 this sequence version replaced gi:1397827.

Contact: Robert Strausberg, Ph.D.

Tel: (301) 496-1550

Email: Robert_Strausberg@nih.gov

Tissue Procurement: Lee Helman, M.D., Michael R. Emmert-Buck, M.D., Ph.D.

cDNA Library Preparation: David B. Krizman, Ph.D.

DNA Sequencing by: Greg Lennon, Ph.D.

Clone distribution: NCI-CGAP clone distribution information can be found through the I.M.A.G.E. Consortium/LNL at: www.bio.llnl.gov/abbr/image/image.html

www.bio.llnl.gov/abbr/image/image.html

Insert Length: 380

Std Error: 0.00

Seq primer: -40ml3 fwd. ET from Amersham.

FEATURES

```

ser: -40ml3 fud. ET from Amer sham.
Location/Qualifiers
1. . 334
/organism="Homo sapiens"
/db_xref="taxon:3606"
/clone="IMAGE:310314"
/collection="NCI-60"
/tissue_type="Ewing's sarcoma"
/lab_host="DH10B"
/note="Vector: pAMP10; mRNA made from Ewing's sarcoma,
cDNA made by oligo-RT priming. Non-directionally cloned.
Size-selected on agarose gel, average insert size 600 bp.
Reference: Krizman et al. (1996) Cancer research
```

BASE COUNT	86 a	73 c	109 t	1 others
ORIGIN				
Query Match	1.28;	Score 36;	DB 34;	Length 334;
Best Local Similarity	50.34;	Pred. No. 10;		
Matches	87;	Conservative 0;	Mismatches 86;	Indels 0;
Gaps	0;			
1446	tcctctatctacttgagctgggagccttcagcagccgggtatcaacggtccattc	atc	1505	
149	TCCTCTCCCAACCGGGGTATGAAGATTATCAGCAGCTGAACATATAAGCCATTC	ATC	208	
1506	gacagtgatctgcagcgggcttcgtcaacttcggcagcgtctcactcttcagatgata	1565		
209	CAAAAAGTGAACACAGAAGAGACTTAAGTCATCTCCTTGGGTTTCTACTTCTACTCTT	268		
1566	ggttcgcgcgcgcagctggccagcgcgcctcgtgttaccaaaagtgcagggcgca	1618		
269	GNAGGGCTCCACTGCTCCCATCTATCTGTCATGTAACCAACATGCGCATTCGA	321		

RESULT 15

NW321625 446 bp mRNA EST 26-JAN-2000
 uc074e09.yt NCI-CGAP MamI Mus musculus cDNA clone IMAGE:2848296 5'
 similar to gb:D78135 Mouse mRNA for CIRP (MOUSE);, mRNA sequence.
 NW323625
 NW323625.1 GI:6757650
 EST.
 house mouse.
 Mus musculus
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
 1 (bases 1 to 446)
 NCI-CGAP <http://www.ncbi.nlm.nih.gov/ncicgap>.
 National Cancer Institute, Cancer Genome Anatomy Project (CGAP).
 Tumor Gene Index
 Unpublished (1997)
 On Jun 22, 1998 this sequence version replaced gi:3246932.
 Other ESTs: uc074e09.xl
 Contact: Robert Strausberg, Ph.D.
 Tel.: (301) 496-1550
 Email: Robert_Strausberg@nih.gov
 Tissue Procurement: Gilbert Smith, Ph.D.
 cDNA Library Preparation: Life Technologies, Inc.
 DNA Sequencing by: Washington University Genome Sequencing Center.
 Clone Distribution: NCI-CGAP clone distribution information can be
 found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html

847820:IGM
MGI:1028748

Seq primer: -40RP from GIBCO
High quality sequence stop: 435.
Location/Qualifiers
1. 446
/organism="Mus musculus"
/strain="FVB/N"
/db_xref="taxon:10090"

FEATURES

source

```

/clone="IMAGE:2648296"
/clone.lib="NCI CGAP.Maml"
/tissue.type="tumor biopsy sample"
/dev.stage="3 months, virgin"
/lab.host="DH108"
/note="Organ: mammary; Vector: pCMV-Sport6; Site.1: Salt
Site.2: NotI; Cloned unidirectionally. Primer: Oligo dt.
Library constructed by Life Technologies. Investigator
providing samples: Gilbert Smith, NIH"
89 a 106 g 164 g 87 t

```

Query Match	1.28;	Score 36;	DB 80;	Length 446;
Best Local Similarity	50.68;	Pred. No. 12;	Indels 0;	Gaps 0;
Matches 8;	Conservative	0;	Mismatches	0;
QY 414	aaggttaagttccacacggaaggcatatcggtctccaaagatctctgggttaaaatatggcc	473		
DB 243	ATGGGAAGTCTGGACGGGGGGGAGATCAGATTGACGAGCTGGCAAGTCTCTGACA	302		
QY 474	agggctcgcatcgctacacgcgttgggtctgtgattggccgcacacctgttggcatcgac	533		
DB 303	ACCGGTCCCGAGATACCGGGGTGGCTCTGCTGGAGCCGGGGCTTTTTCGTGGGGGAC	362		
QY 534	gaggtcacgcgcagtggtgttttccctaggaaacacaaagcacacccggggagg	585		
DB 363	GAAGCCGGGGCCGAGGGTCTCCAGAGAGGAGGAGACCCGGGGTATGGAGG	414		

Search completed: May 17, 2000, 05:14:39
Job time: 2905 sec

Thu May 18 22:21:30 2000

us-09-151-189-2_copy_441_676.441-676.rge

Page 1

OM of: US-09-151-189-2_COPY_441_676 to: GenEmbl:* out_format: pfs
Date: May 17, 2000 10:56 AM
About: Results were produced by the GenCore software, version 4.5,
Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:
-MODEL-frame_r2n_model -DEV-rlp
-O/cgn2_1/USPTO_spool/US09151189/runat_15052000_082513_1608/app_query.fasta.1
-DB-GenEmbl -OFMT-fastcap -SUFFIX-441-676.rge -GAPOF-12.000
-GAPEXT-4.000 -MINMATCH-0.100 -LOOPCL-0.000 -LOOPEXT-0.000
-GAPOF-4.500 -GAPEXT-0.050 -XGAPOF-10.000 -XGAPEXT-0.500
-FGAPOF-6.000 -DELEXT-7.000 -XGAPOF-10.000 -XGAPEXT-0.500
-DELOF-6.000 -DELEXT-7.000 -START-1 -MATRIX-blosum62
-TRANS-human40.cdi -LIST-45 -DOALIGN-200 -THR_SCORE-pct
-ALIGN-15 -MODS-LOCAL -OUTFMT-pfs -NORM-ext -MINLEN-0
-MAXLEN-1000000 -USER-US09151189 -NCPU-6 -ICPU-3 -NO_XLPY -WAIT
-THREADS-1

Search information block:
Query: US-09-151-189-2_COPY_441_676
Query length: 236
Database: GenEmbl:*
Database sequences: 882769
Database length: -486395729
Search time (sec): 483.650000

Sequence	Strd Orig	zScore	Escore	Len	Documentation
gb_p12.AF053411	+	1238.00	2156.07	4,9e+112	2931
gb_p12.D87657	+	327.50	556.99	5,7e+23	2035
gb_p12.E17199	+	325.50	554.68	7,7e+23	1794
gb_p12.E17200	+	318.00	541.49	4,2e+22	1791
gb_p12.D87658	+	318.00	540.29	4,2e+22	2029
gb_p12.AF001953	+	116.50	158.33	0,25117	12198
gb_p12.NR0026	-	103.50	134.11	10,76	23740
gb_p12.TC001272	-	98.00	132.72	1,88	2103
gb_p12.TC006070	-	95.00	127.86	3,50	2013
gb_p12.AF044733	-	94.00	138.83	11,16	4292
gb_p12.AF044732	-	94.00	135.52	17,06	6055
gb_p12.Y098H12	-	92.00	104.72	886,92	103576
gb_p12.AF001696	-	90.50	120.32	119,86	15513
gb_p12.ACPNPDNA	+	87.50	122.29	93,13	7297
gb_p12.AB000134	+	87.00	141.42	8,00	909
gb_p12.AF193495	+	87.00	141.25	8,19	926
gb_p12.AB000133	+	87.00	137.82	12,71	1323
gb_p12.AB001946	+	87.00	109.10	505,37	26263
gb_p12.D87659	+	86.50	140.28	9,26	934
gb_p12.RE282004	+	86.50	121.52	102,80	6583
gb_p12.A48542	+	86.50	92.57	4,2e+03	133894
gb_p12.L22858	+	86.50	92.57	4,2e+03	133894
gb_p12.AC010545	+	86.50	87.22	4,2e+03	233554
gb_p12.AC038024	+	86.00	131.98	26,86	2022
gb_p12.AC038023	+	86.00	131.86	27,29	2048
gb_p12.WNCEP1	-	86.00	123.67	76,05	4704
gb_p12.MUSEG1	-	86.00	123.78	76,95	4749
gb_p12.I08331	-	86.00	123.77	76,97	4750
gb_p12.AC012251	-	86.00	89.49	6,3e+03	168280
gb_p12.AC012251	+	85.50	132.06	26,59	1830
gb_p12.CHRNAVPO	+	85.50	124.23	72,63	4135
gb_p12.HSM01979	+	85.50	114.97	237,96	10830
gb_p12.AC001964	+	85.50	111.66	364,12	15294
gb_p12.AC0000753	+	85.00	125.51	61,60	3301
gb_p12.RE282005	+	84.00	94.31	3,4e+03	70680
gb_p12.AC023415	+	84.00	82.96	1,4e+04	230278
gb_p12.NC068299	+	83.50	83.05	1,4e+04	2882
gb_p12.AC031631	+	83.50	83.05	1,4e+04	2882
gb_p12.HSA324122	+	83.00	134.62	19,15	887
gb_p12.ECU19784	+	83.00	132.20	26,12	1041
gb_p12.S76749	+	82.50	120.45	117,93	3537
gb_p12.HSM800425	-	82.50	120.45	117,93	3537

gb_bai:CLOBONT + 82.50 118.80 145.70 4199 L35456 Clostridium botulinu
gb_bai:CBONTEG + 82.50 118.78 146.13 4209 X81714 C. botulinum bont/f 9
gb_bai:D67490 - 82.50 109.38 487.56 11188 U67490 Methanococcus janna
gb_bai:AE001898 - 82.50 108.30 560.23 12523 AE001898 Deinococcus radio
seq_name: gb_p12.AF053411

seq documentation block:
LOCUS AF053411 2931 bp mRNA PLN 12-SEP-1998
DEFINITION Fucus distichus vanadium bromoperoxidase mRNA, complete cds.
ACCESSION AF053411
VERSION AF053411.1 GI:3582762
KEYWORDS
SOURCE Fucus distichus.
ORGANISM Fucus distichus.
Sukaryota; stramenopiles; Phaeophyceae/Xanthophyceae group;
Phaeophyceae; Fucales; Fucaceae; Fucus.
REFERENCE 1 (bases 1 to 2931)
AUTHORS Vreeland V., Ng K. and Epstein L.
TITLE Direct Submission
JOURNAL Submitted (11-MAR-1998) Environmental Science, Policy and
Management, University of California, 201 Wellman Hall, Berkeley,
CA 94720-3112, USA
FEATURES
Location/Qualifiers
1..2931
/organism="Fucus distichus"
/db_xref="taxon:3012"
/dev_stage="2 cell embryo"
1..227
228..2258
/cdoon_start=1
/product="vanadium bromoperoxidase"
/db_xref="gi:3582763"
/translation="MLCHRAADTRGSPMDTGVLRLLTSEORAKWRQLEGEKSLGF
HPESTPYKLEGETKWKVKLPDTSISAKILGKIMARVRIATLAVVLAAPCLAFD
EYTAGPYKLEGETKWKVKLPDTSISAKILGKIMARVRIATLAVVLAAPCLAFD
GVCTGTTGDTLAFMKFSLPHDELGOVTADEEDLNGDSTCEVDVAGDPA
GLVPTAFAFADLSGPAFATIPPPVTLSPSPALAOALYLMALARDVPPVQYGT
DRITTAANLAGMGFPNDVSGSDTVPFSQLFRATFVGTEGPFVSQLVNS
VFQVNTATYRGLSLLELGAFSRPPINGPFIQSDRQAGFVNFETSHVRLIGAA
LAQRASCKQKQVHRFARPAALGCTLHNTIAGDADFISLENDLLKRVAPINAA
QPNNEVYLLPOAVGSPTHPSYSGHATONGAFATVLAIGLDGSGGCPNPVF
PDDGLELNFEGACLTVEGINKLVANVAFGRMLGIHIFDCIGLIGETITVFT
LHQLMTFAEATEFRLFTGEVTKLFQDGTFSIDGDCSGLVITGVADCOA"

BASE COUNT 694 a 776 c 784 g 677 t
ORIGIN
3'UTR
CDS
5'UTR

alignment_scores:
Quality: 1238.00 Length: 236
Ratio: 5.245 Gaps: 0
Percent Similarity: 100.000 Percent Identity: 100.000
alignment_block:
US-09-151-189-2_COPY_441_676 x AF053411
Align seg 1/1 to: AF053411 from: 1 to: 2931
1 HistyrPheArgLeuLeuGlyAlaAlaGluLeuAlaGlnArgAlaSerCy 17
1548 CACTACTTCAGATTGATAGTCCGCGAGCTGGCGAGCGTCTCGCTG 1597
17 styrcLnysTrcglNvalHisarPhealaArgProclalaLeuGlyg 34
1598 TTACCAAAAGTGGCAGTGCATGATTCGACGCCCGCGAGCTCTCGGG 1647
34 lYThrLeuHisAsnThrIleAlaGlyAspLeuAspAlaAspPheAspIle 50
1648 GTNCCCTCCACACACACCTCCGCGGGGATCTAGATGCGACACTCGACATC 1697

137 PheGluGluValAlaCysLeuThrTyrGluGluGluLeuLeuAlaVala 153
 1703 TTCAAGGGAAGT...CTCACTGTGGCGGTGAATGACAAAGCTCGCGA 1749
 153 IasnValAlaPheGluGluMetLeuGluGluLeuLeuLeuLeuLeu 170
 1750 CAATATTCGGATCGGGGTACATGCGAGCTGTCTCACTACTCTCTGAC 1799
 170 LylleGluGluLeuLeuGluGluGluGluGluGluGluGluGluGlu 186
 1800 AGTTGAGTCACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1849
 187 GlnGluLeuMetThrPheAlaGluGluAlaThrPheGluPheArgLeu 203
 1850 GAGCAAGTCTCAGTATGCGGAGAGTCTTCTTCACTTTCGCGAAGTT 1899
 203 eThrGluGluValTleLysLeu 210
 1900 TGATGGAAGTCAATCCAGATC 1921

seq_name: gb_pat:EL1799
 seq_documentation_block: 1794 bp DNA PAT 28-JUL-1999
 LOCUS E17199
 DEFINITION Corallina pilulifera mRNA for haloperoxidase.
 ACCESSION E17199
 VERSION E17199.1 GI:5711882
 KEYWORDS JP 1998248581-A/1.
 SOURCE
 ORGANISM
 unclassified.
 unclassified.
 unclassified.
 IZUMI, Y. and Tanabe, T.
 NEW HALOPEXIDASE GENE AND ITS UTILIZATION
 Patent: JP 1998248581-A 22-SEP-1998;
 OTSUKA PHARMACEUT CO LTD
 OS Corallina pilulifera
 PN JP 1998248581-A/1
 PD 22-SEP-1998
 PF 06-MAR-1997 JP 1997070539
 PI IZUMI YOSHIKAZU, TANABE TADASHI
 PC CL2N15/09, CL2N8/08, (CL2N15/09, CL2N1/89), (CL2N8/08, CL2N1/19);
 CC strandedness: Double;
 CC topology: Linear;
 FH Key
 FT Location/Qualifiers
 FT source 1..1794
 FT /clone='BP01' /organism='Corallina pilulifera' FT
 FT mat_peptide 1..1794
 FT Location/Qualifiers
 FT 1..1794
 FT /organism='unidentified' /db_xref='taxon:32644'
 BASE COUNT 395 a 481 c 505 g 413 t
 ORIGIN

alignment_scores:
 Quality: 325.50 Length: 224
 Ratio: 2.276 Gaps: 8
 Percent Similarity: 63.839 Percent Identity: 38.393

alignment_block:
 US-09-151-189-2_COPY_441_676 x E17199
 Align seg 1/1 to: E17199 from: 1 to: 1794

1 HistyrPheArgLeuLeu...GlyAlaGluLeuAlaGlnArgAlaSe 16
 1138 CACGTCGTGAGTCTGTGAGTGAAGTGGTACGCGCGGTGAAGCGGT 1187

16 ICySTyTGlnLysTyrGlnValHisATGPhoAlaATGPTGlnAlaLeuG 33
 1188 ACGTACAGAGAGTTTACATTCATCGTCGCTCGCCCTGAGGCTACCG 1237
 33 LgLy...ThrLeuHisThrIleAla...GlyAspLeu 44
 1338 GTGTCGTGATTCCTTACAAATCCACCGCAGAGGCCGAGAGCATT 1287
 45 AspAlaAspPheAspIleSerLeuLeuGluAsnAspGluLeuLeuLeu 61
 1288 TTCCTGAGGTGATTCCTTGTGTAAGAGCTTGGAGATATCTTGGAA 1337
 61 gValAlaGluLeuAsnAlaIleGlnAsn...ProAsn 73
 1338 AGCTGAATTAACATAGGACAGACATAGCTGACGGAGATCTCTGACC 1387
 73 snGluValThrTyrLeuLeuProGlnAlaIleGlnValGlySerProThr 89
 1388 CTGATCCTTCATTCTCTGTCGATGCAATTCGCGAGGCGAGCCCATTC 1437
 90 HisProSerTyrProSerGlyHisAlaThrGlnAsnGlyAlaPheAla 106
 1438 CATCCCTCTACGAGCGGCCACGCTGTGTTGCTGGCGCATGTGTGAC 1487
 106 KValLeuLysAlaLeuLeuGluLeuAspArgGlyGlyGluCysPheProA 123
 1488 GATCCCGAGGCGGTC...TTCGACTCCGCGCATCGAGATC...G 1525
 123 snProValPheProSerAsp...AspGlyLeuGluLeuLeuLeu 136
 1526 ATCAGGTGTTCAGTTCGACGACAGATGAGGACAGCTGTGAGTCTGCT 1575
 137 PheGluGlyAlaCysLeuThrTyrGluGluGluLeuLeuLeuAlaVal 153
 1576 TTCAGGAGACT...CTCAGTGTTCGCGGTGAATGAACAACTCGCGA 1622
 153 IasnValAlaPheGluGluGlnMetLeuGluGluLeuHisTyrArgPheAspG 170
 1623 CAATATTCGAGTCGGCGCTAATACATGCGAGGTGTCTACTACTCTCTGACC 1672
 170 LylleGluGluLeuLeuGluGluGluGluGluGluGluGluGluGlu 186
 1673 AGTTCAGTCACTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 1722
 187 GlnGluLeuMetThrPheAlaGluGluAlaThrPheGluPheArgLeu 203
 1723 GAGCAAGTCTCAGTATGCGGAGAGTCTTCTTCACTTTCGCGAAGTT 1772
 203 eThrGluGluValTleLysLeu 210
 1773 TGATGGAAGTCAATCCAGATC 1794

seq_name: gb_pat:EL17200
 seq_documentation_block: 1791 bp DNA PAT 28-JUL-1999
 LOCUS E17200
 DEFINITION Corallina pilulifera mRNA for haloperoxidase.
 ACCESSION E17200
 VERSION E17200.1 GI:5711883
 KEYWORDS JP 1998248581-A/2.
 SOURCE
 ORGANISM
 unclassified.
 unclassified.
 unclassified.
 IZUMI, Y. and Tanabe, T.
 NEW HALOPEXIDASE GENE AND ITS UTILIZATION
 Patent: JP 1998248581-A 22-SEP-1998;
 OTSUKA PHARMACEUT CO LTD
 OS Corallina pilulifera
 PN JP 1998248581-A/2
 PD 22-SEP-1998
 PF 06-MAR-1997 JP 1997070539

Align seg 1/1 to: D87658 from: 1 to: 2029

```

1 HistHphAargLeuile...GlyAlaAlaGluLeuAlaGlnAargLase 16
1250 CAGGCTGAGTCTGCTGACTGAGTGGCTAGCGCGGCTTGAGCGCGT 1299
16 rCysTyrGlnLysTrpGlnValHisArgPheAlaArgProGluAlaLeu 33
1300 AGGATACAGAGATTACATTCATCCCTGCGCGCTGAGGCTAGT 1349
33 lCysLysThrLeu.....HisAsnThrIleAlaGlyAspLeuAsp 45
1350 GTGGTCTGATTAGCGTTAACAGAAATCTTCTTCGCGGTAGT..... 1393
46 AlaAspPheAspIleSerLeuLeuGluAsnAspGluLeuLeuLysArg 61
1394 .....GATATAATATTCTGAGTTAGTGAATCTGTCGAGGAGCT 1434
62 .....ValAlaGluLeuAsnAlaAlaGlnAsnProA 72
1435 CTATCAATTCGGATGACGTTCTGAGAGCAATGAAACAGACAGGG 1484
72 snAsnGluVal.....ThrTyrLeuLeuProGluAlaIle 83
1485 CTGACGGGATTGTGAGCGCGGATTAATCATCTCTGCGGATGCGATT 1534
84 GlnValGlySerProThrHisProSerTyrProSerGlyrAlaThrG1 100
1535 GCGAGGGCAGCCCATTCATCCATCCCTCTATGGAAGTGGCCAGCGTGGT 1584
100 nAsnGlyAlaPheAlaThrValLeuLysAlaLeuLeuGly..... 113
1585 TGCTGCGCATGTGTGCAATCTCTCAAGCGTCTTCGACGCCAATCTTC 1634
114 ..LeuAspArgGlyLeuLysCysPheProAsnProValPheProSerAsp 129
1635 AGATGATAAAGGTGTTCAG.....GTGACACTGATGAG 1669
130 AspGlyLeuLeuLeuLeuLeuAsnPheGluGlyAlaCysLeuTyrGluG1 146
1670 GACAGCTGTGAGTGTCTCTTCAGGGACT...CTCATCTGTGCGG 1716
146 yGluLeuAsnLysLeuAlaValAsnValAlaPheGlyArgGlnMetLeuG 163
1717 TGAATGAAACAGCTCGCCGCAATGTGGATCGGCGGCAACATGCGGG 1766
163 lylleHisTyrArgPheAspGlyIleGlnGlyLeuLeuLeuLysLeuThr 179
1767 GTGTTCACTACTCTCTGACAGTTCGAGTCTGAGTCTGCTGCTGTCG 1816
180 IleThrValArgThrLeuHisGlnGluLeuMetThrPheAlaGluGluAl 196
1817 ATTGCGATTGGAATCTTGAGGACGACGCTGACGATGCGGAGAACTT 1866
196 aThrPheGluPheArgLeuPheThrGlyGluValIleLysLeu 210
1867 CTCTTCATTTGCGAATTCGACGAACTACATCCAGATT 1909
seq_name: gb_ba2:AE001953
seq_documentation_block:
LOCUS AE001953 12198 bp DNA BCT 22-NOV-1999
DEFINITION Deinococcus radiodurans RI section 90 of 229 of the complete
chromosome 1.
ACCESSION AE001953 AE000513
VERSION AE001953.1 GI:6458740
KEYWORDS
SOURCE
ORGANISM Deinococcus radiodurans
Bacteria; Thermus/Deinococcus group; Deinococcus
REFERENCE
1 (bases 1 to 12198)
AUTHORS White, O. Eisen, J.A., Heidelberg, J.F., Nicky, E.K., Peterson, J.D.,
Dodson, R.J., Haft, D.H., Gwinn, M.L., Nelson, W.C., Richardson, D.L.,
Moffat, K.S., Qian, H., Jiang, L., Pamphile, W., Crosby, M., Shen, M.,
Fleischmann, R.D., Ketchum, K.A., Nelson, K.E., Salzberg, S.,
Smith, H.O., Venter, J.C. and Fraser, C.M.
Genome sequence of the radioresistant bacterium Deinococcus
radiodurans RI
Science 286 (5444), 1571-1577 (1999)
2 (bases 1 to 12198)
White, O., Eisen, J.A., Heidelberg, J.F., Hickey, E.K., Peterson, J.D.,
Dodson, R.J., Haft, D.H., Gwinn, M.L., Nelson, W.C., Richardson, D.L.,
Moffat, K.S., Qian, H., Jiang, L., Pamphile, W., Crosby, M., Shen, M.,
Fleischmann, R.D., Ketchum, K.A., Nelson, K.E., Salzberg, S.,
Smith, H.O., Venter, J.C. and Fraser, C.M.
Genome sequence of the radioresistant bacterium Deinococcus
radiodurans RI
Science 286 (5444), 1571-1577 (1999)
20036896
TITLE
JOURNAL
MEDLINE
REFERENCE
AUTHORS
FEATURES
source
1 12198
/organism="Deinococcus radiodurans"
/strain="RI"
/db_xref="taxon:1299"
/chromosome="1"
92..997
/gene="DRI011"
92..997
/gene="DRI011"
/feature="note"
/note="similar to GB:AE000657 percent identity: 62.08;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=1
/product="conserved hypothetical protein"
/protein_id="AF10587.1"
/db_xref="GI:6458741"
/translation="MTSSAPSSSLPDVHLHGYSCFENDFIFHAIHAGLVGGLPVPQ
EYEDVQTLNPNVAGRPWPKISYAFVDMRYVALRSGALGCGVPLVTRPGV
KDTALRVASPGALITAEILLAIPTGVVVRMRIDVPAVGRTGIDGTADAGLI
HESRPTTGLDKLLDCAWMEGETGLPLGAILVRDLPAQTRQLQDQVRRSLE
YAYAPAAADYIRHAELEMDVQAHIDLYNSLSDVGEGERAVRELLURRAVEV
GAVTARPEPLFVGDAEAAGTPTL"
1033..1956
/gene="DRI012"
1033..1956
/gene="DRI012"
/note="similar to GB:AL009126 percent identity: 54.74;
identified by sequence similarity; putative"
/codon_start=1
/transl_table=1
/product="ABC transporter, ATP-binding protein"
/protein_id="AF10588.1"
/db_xref="GI:6458742"
/translation="MSKASAPAPRAVEVGGISKSGYGRANVLTDFVNLVIGEVALTG
PNEGATTLRLTGLAETPEVRIACVYVHNGCPARALMGAVVAPAKYPTPTG
TENKIHASLIGAGVCPGRISEVLALELSMGRRVAEYSGGRIQVGSARILAE
KYLIDEPISGLDPLGIGLIRHIVTSLATSCCAVILSTHREIATYANTVGTGGR
LVTVDRSTSAIRFVRDPPVGAALLERQPIYVRKATPRPPIYVAQLGAQANVPEAL
RLANEDVAVTEVTRFDLYEYRVEYA"
1956..2729
/gene="DRI013"
1956..2729
/gene="DRI013"
/note="identified by Glimmer2; putative"
/codon_start=1
/transl_table=1
/product="hypothetical protein"
/protein_id="AF10594.1"
/db_xref="GI:6458748"
/translation="MLSLISLFEFRKLASRSARIALVCTLLPFLMSAPRLGELRV
OLVSCQLPAVSLGAVGAPLPLTAVAVAEIGVAECVACVACVAPLAPLAPSRVIVS
KLIVALLPLLVLLATLLGSLVAGIPHGVEFGGTGGLGGLFVGGLVSPSAAIAEV

```


alignment_scores:
 Quality: 95.00 Length: 243
 Ratio: 0.905 Gaps: 11
 Percent Similarity: 43.210 Percent Identity: 22.222

alignment_block:

US-09-151-189-2_COPY_441_676 x TC006070/rev.

Align seq 1/1 to reverse of: TC006070 from: 1 to: 2013

```

1 HisTyrPheArgLeuLeuGlyAlaAlaGluLeuAlaGlnArgAlaSerCys 17
1282 CATGCTTCCTAGTCTGGGACGACATCTCTGGGAG...GGCATATG 1236
17 sTyrGlnLysTrpGlnValHisArgPheAlaArgProGluAla..... 31
1235 TCACAAAGATGG.....ACGGGCGCCAGGCGCGCTCTT 1201
32 .....LeuGlyGlyThrLeuHisAsnThrIleAlaGlyAspLeu 44
1200 CCATGAAGCTCTGTCGACGCGCTCGCGCACTGT..... 1163
45 AspAlaAspPheAspIleSerLeuLeuGluAsnAspGluLeuLeuLysAr 61
1162 .....CTGCTTAAAG 1152
61 gValAlaGluIleAsnAlaAlaGlnAsnProAsnAsnGluValThrTyrL 78
1151 A.....ACAAAGCCAAA..... 1139
78 euLeuProGlnAlaIleGlnValGlySerProThrHisProSerTyrPro 94
1138 .....GCACCGGGATACCTAGTGGCCCG 1115
95 SerGlyHisAlaThrGlnAsnGlyAlaPheAlaThrValLeuLysAlaLe 111
1114 GTCATCATACATCTGACGAAACATCTGCGCACGCGCTCAACACGCT 1065
111 uLeuGlyLeuAspArgGlyGlyGlyCysPheProAsnProValPhePro 128
1064 C..... 1064
128 erAspAspGlyLeuGluLeuIleAsnPheGluGlyAlaCysLeuThrTyr 144
1063 .....CTTCAGCGCGCTCGTGTGACAGATATCTCAAGTCTT 1025
145 GluGlyGluIleAsnLysLeuAlaValAsnValAlaPheGlyArgGluMe 161
1024 GAGCTAAGGTTCGGGACTCTGTG..... 998
161 tLeuGlyIleHisTyrArgPheAspGlyIleGlnGlyLeuLeuLeuGly 178
997 .....TTGTGTG..... 989
178 luThrIleThrValArgThrLeuHisGlnGluLeuMetThrPhe..... 192
988 ..ACGCTCGCGCGCGACAGCTCTGCGCTCGCGTG...ACGTTTCATCTG 944
193 ...AlaGluGluAlaThrPheGluPheArgLeuPheThrGlyGluValI 208
943 AGAGCGAAAGCGGACACACTCTTCATCTCTCTTCCTTAGCCAATGGAC 894
208 elysLeuPheGlnAspGlyThrPheSerIleAspGlyAspMetCysSerG 225
893 TCCTCTCTCTGCTGAGTCTCTCGACAGCTTGGCGGCGCTCTGATG 844
225 yLeuValTyrThrGlyValAlaAspCys. 234
843 GCTACGCAACGACGCAATCCGACGTGT 815

```

seq_name: gb_inl:AF044733

seq_documentation_block:

LOCUS AF044733 4292 bp DNA INV 11-FEB-1998
 DEFINITION Trypanosoma cruzi clone Blk poly(A)-binding protein (PABP) gene, complete cds.
 ACCESSION AF044733
 VERSION AF044733.1 GI:2854067
 KEYWORDS
 SOURCE Trypanosoma cruzi #
 ORGANISM Trypanosoma cruzi
 Eukaryota; Euglenozoa; Kinetoplastida; Trypanosomatidae;
 Trypanosoma; Schizotrypanum.
 REFERENCE 1 (Bases 1 to 4292)
 AUTHORS Batista, J.A.N. and Martins de Sa, C.
 TITLE Direct Submission
 JOURNAL Submitted (26-JAN-1998) Biologia Celular, Universidade de Brasilia, Campus Universitario - Asa Norte, Brasilia, DF 70910-900, Brazil
 FEATURES
 Location/Qualifiers
 1..4292
 /organism="Trypanosoma cruzi"
 /strain="dm 28c"
 /db_xref="taxon:5693"
 /clone="Blk"
 complement(781..1102)
 /note="RIME-like element"
 <781..>4292
 /gene="PABP"
 1769..1797
 /gene="PABP"
 /note="trans splicing signal"
 1825..1826
 /gene="PABP"
 /note="secondary spliced leader acceptor site"
 1827..>4292
 /gene="PABP"
 1827..2360
 /gene="PABP"
 1828..1873
 /gene="PABP"
 /note="trans splicing signal"
 1879..1880
 /gene="PABP"
 /note="major spliced leader acceptor site"
 2361..4013
 /gene="PABP"
 /codon_start=1
 /product="poly(A)-binding protein"
 /protein_id="AAC02538.1"
 /db_xref="GI:2854068"
 /translation="MSNPAPASPTVGGGLDPLNEKLYDHEVRLGVPASVVRVCVDS
 VTQSLGTVNFQNPADAEXALQDQNGVKLSKHRIKIQDRPSKRGVNTIVK
 LPSSVDIALKEFSTYGLTAIGLAVREKSGKATSTIEESALCAQVQVAGV
 LDDCALIVRIQPHREGLKQFINLYKLDPSYNDERLKEVFSAGFEVTSKAYRDL
 GANQVGFATVATHEAAKAVELEKELPERGCTETSVSYMDRATRVSREREL
 RREQQGHSEPNLYKFPDDTTSERLKEFRCGTSVSYMDRATRVSREREL
 SFKEQSAASATQELGSGTALGPPLFVTVLAKRDAROTLEDRKQPRMQRPGV
 LMCGMGPGFGFQAFNGVYFVNFPMKMTPTPMGSGQLPFGMPTPMQVRA
 PPRPQPIAPPQPOPLASQGNLSTVLANTPEQKNVLERLYNHIVAINPAA
 AKVTGLEMDNGEILNLDTPGLDARYOEALVLRHNMV"
 4014..>4292
 /gene="PABP"
 3'UTR
 BASE COUNT 1183 a 958 c 1056 g 1093 t 2 others
 ORIGIN
 alignment_scores:
 Quality: 94.00 Length: 243
 Ratio: 0.895 Gaps: 11
 Percent Similarity: 43.210 Percent Identity: 22.222
 alignment_block:
 US-09-151-189-2_COPY_441_676 x AF044733/rev


```

17  STYGLNLTSTPGLNVALHISAGPHEALAAAGPAGGGLAALA..... 31
|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1729 TCACAAAAGTGG.....ACGCGCCCAAGCGCGTGT 1695
32  .....LeuGlyGlyThrLeuHisAsnThrIleAlaGlyAspLeu 44
|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1694 CCATGAAGCTCGTGCATGACGCGCTGCGGCACTCTG..... 1657
45  ASPALAASPHEASPTLSERLEULEUGLUASNAPGLULEULEULYSAR 61
|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1656 .....CTCTTAAAG 1646
61  GVALALAGLUILEASNAALAGLNASNPROASAGLUVALTHRTYR 78
|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1645 A.....ACAAGGCCAA..... 1633
78  euLeuProGlnAlaIleGlnValGlySerProThrHisProSerTyPro 94
|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1632 .....GCACCGGATACCTAGTGGCCG 1609
95  SerGlyHisAlaThrGlnAsnGlyAlaPheAlaThrValLeuAlaLe 111
|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1608 GTCCATCATACACTGACGAACAGTCTCGCCACAGCGCTCAACAGCT 1559
111 uileGlyLeuAspArgGlyGlyCysPheProAsnProValPhePro 128
|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1558 ..... 1558
128 erAspAspGlyLeuGluLeuIleAsnPheGluGlyAlaCysLeuThrTy 144
|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1557 .....CTCAGCGCTGCTGTCGACAGTATCTCAAGTCTT 1519
145 GluGlyGluIleAsnGlyLeuAlaValAsnValAlaPheGlyArgGln 161
|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1518 GAGCTAAGGTCGGGTACTGCTG..... 1492
161 tLeuGlyTleAsTyraGpPheAspGlyIleGlnGlyLeuLeuLeuGly 178
|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1491 .....TTGTGTGTG..... 1483
178 luThrIleThrValArgThrLeuHisGlnGlnLeuMetThrPhe..... 192
|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1482 ...ACGCTCGCGCGAGAGTTCGCTGCGCTG...ACGTCATCTCG 1438
193 ...AlaGluGluAlaThrPheGluPheArgLeuPheThrGlyGluValI 208
|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1437 AGAGCGAGCGCGACACTCAGTTCATCTCTCTTAGCCATGGAC 1388
208 eLysLeuPheGlnAspGlyThrPheSerIleAspGlyAspMetCysSer 225
|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1387 TCCTCTCTGCTGAGTCTCTCGACAGCTTGGCGGACGCTTCGTGATG 1338
225 LysLeuValThrGlyValAlaAspCys 234
|:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||:
1337 GGTACCGGACGTAGCAATCCGACTGT 1309

seq_name: gb_p12:YUP8H12
seq_documentation_block:
LOCUS YUP8H12 103576 bp DNA PLN 29-SEP-1997
DEFINITION Arabidopsis thaliana chromosome 1 YAC YUP8H12 complete sequence.
ACCESSION AC000098
VERSION AC000098.1 GI:2358139
KEYWORDS HTG
SOURCE
ORGANISM Arabidopsis thaliana
Eukaryota; viridiplantae; Streptophyta; Embryophyta; Tracheophyta;
euphyllophytes; Spermatophyta; Magnoliophyta; eudicotyledons; core
eudicots; Rosidae; eurosids II; Brassicales; Brassicaceae;
Arabidopsis.
REFERENCE 1 (bases 1 to 103576)
Theologis A., Osborne B.I., Vysotskaya V.S., Federspiel N.A.,
Toriumi M., Yu G., Ojima O., Araujo R., Chung E., Dewar K.,

```

```

Dietrich P., Ecker J.R., Marziani A., Oefner P. and Davis R.W.
The sequence of YAC YUP8H12 from Arabidopsis thaliana chromosome 1
Unpublished (1997)
We have determined that YAC YUP8H12 is chimeric, and is comprised
of two distinct genomic EcoRI fragments from chromosome 1. This
submission contains the sequence from the EcoRI site at position
181919 to position 285516 (left end) of our previous Phase II
submission. This fragment contains the AXR1 locus and the marker
YUP8H12L.
2 (bases 1 to 103576)
Theologis A.
Direct Submission
Submitted (25-NOV-1996) Plant Gene Expression Center, 800 Buchanan
Street, Albany, CA 94710, USA
3 (bases 1 to 103576)
Theologis A., Davis R.W. and Federspiel N.A.
Direct Submission
Submitted (03-SEP-1997) Plant Gene Expression Center, 800 Buchanan
Street, Albany, CA 94710, USA
4 (bases 1 to 103576)
Theologis A., Davis R.W. and Federspiel N.A.
Direct Submission
Submitted (06-SEP-1997) Plant Gene Expression Center, 800 Buchanan
Street, Albany, CA 94710, USA
5 (bases 1 to 103576)
Theologis A.
Direct Submission
Submitted (10-SEP-1997) Plant Gene Expression Center, 800 Buchanan
Street, Albany, CA 94710, USA
6 (bases 1 to 103576)
Theologis A.
Direct Submission
Submitted (12-SEP-1997) Plant Gene Expression Center, 800 Buchanan
Street, Albany, CA 94710, USA
7 (bases 1 to 103576)
Theologis A.
Direct Submission
Submitted (29-SEP-1997) Plant Gene Expression Center, 800 Buchanan
Street, Albany, CA 94710, USA
On Sep 5, 1997 this sequence version replaced gi:1932823.
FEATURES
Source
1. 103576
/organism="Arabidopsis thaliana"
/cultivar="Columbia"
/db_xref="taxon:3702"
/chromosome="1"
/clone="YUP8H12"
3272..6134
/gene="YUP8H12.1"
join(3272..3379,3486..3503,4168..4256,4507..4564,
4922..5057,5173..5236,5466..6134)
/gene="YUP8H12.1"
/codon_start=1
/evidence="not experimental"
/product="YUP8H12.1"
/protein_id="AAH1441.1"
/db_xref="GI:2388560"
/translation="MGKEQKHSHSVESDDKVEAVLHLKHSPTLKOEFKFNAC
VGFRLKGNVKAQKLRCSLSRSLGIESLADFEATAEALGLAYAGLDDECR
PVLVAISTMRNVEQVILFIDAIAAVLVLTLYLVSFLNVLPLLCSTAHLPFF
FFETIKPKLALISISFEHTCTQNSRGITFYDLSTATMIVSMQDQSDYD
DFSSTYPSVSSLRFRTSLKTDKIGSCASSRFAFTVSRDGLDVKPCLLTDS
TKLHNTGAYISPLNARSFSPARSEPGGPRGFASFTPMPTDRISITLR
DPRIPSTFQSPAIFFRESRSKSRPRDSFVQFLKFRYRPYDENTYRSKMRPLG
GLVSTQTRRRRHSLSQRF"
complement(8274..10736)
/gene="YUP8H12.2"
complement(join(8274..8453,8547..8726,8824..8940,
9062..9210,9447..9734,9852..9972,10189..10400,
10691..10736))
/gene="YUP8H12.2"
note="Similar to Arabidopsis hypothetical protein
PID:e326839 (gb|297337)"

```


/citation="(3)
 /replace="at"
 3962
 /citation="(3)
 /replace="ct"
 complement(4188..4730)
 /note="ORF 540"
 /codon_start=1
 /protein_id="CAA50545.1"
 /db_xref="GI:525108"
 /db_xref="SWISS-PROT:Q06591"
 /translation="KASMSNIPDILYVNAQINSEEDENVLDIFIEDEYILAKRALAAHI
 IKVAFSGURLLYKRAYTSCGNGYGLVNLNGEYIDINAIFNCASIKLNGOMLF
 QTKINSDSKSTDAARVTSSTRTVETENDDDGEAAASAAIDQEGNADVVGLDFEEN
 IDGDAPTPKKRLDNKOD"
 complement(4586..4729)
 /note="ORF 141"
 4695..5507
 /note="ORF 810"
 4696..7239
 /note="ORF 2541"
 /codon_start=1
 /protein_id="CAA50547.1"
 /db_xref="GI:1185073"
 /db_xref="SWISS-PROT:Q06570"
 /translation="WNSGVMLMAEALILIAFLATLALIEFEDSTTEKLOVMEY
 VKTNADSTPDVIGYVSDIMONYIYVWTVYDLSYHESVDDRIELIFDLQKQFQ
 LVDRIYHVRANDENEFILSGDAVAKPCPAIFNDIAQKCYVPPCDKNSAG
 LYPDELRLDTLYLAQHLDKDI STNAHLIETPILRCFANGRAVERCPDNTFDAET
 GCKVNEZERNPDGTLISIPFSLVYVQVNGRNVVGGCPANKRIFDRNLMSQVE
 AHCPCAGAGHTITADIGDTQIFKLNNESSQITLNLKNSDNQTECDDSRDID
 LPNGTCQVKKVYDODISYNSGQVLCDFEVSIDIECSNVFNALFMDRFLRNF
 PTVFQGTCPATADNVNFIKSTFAENIPNHYGIDMTMLGTFKVLKSKDLS
 LNDIAFAPQLLYARDKDTIGLNFTEPCDFGLDVLDFDARANTCDSGTSVLK
 TLNPGDEGLHLSLTKKDEYDQFCALSYENQKIYENEFORILTILOSDVC
 ADLYTTLQYTLKSKYITPPLATNHLKRNKTEIYGANTRLNKATIPKNAATIP
 PVNFTFNPNRNDLSILFNFQTDVAVWISPEGGDDHVVVPTAPPPPEPE
 PEPEPEPEPEPLDNLKDLFYCHTSVFPFKLTSCHAEVDVI DALNELRNV
 KDADCEGLARLDSHLYNAVYVNGIGCRSAVDGDAIVVYKKAEPVSHVYALNTQSD
 GVYNRLHVRNGQVMACPELYDNNEFKCNIESDKLYLDNLQDSIV"
 complement(4753..4956)
 /note="ORF 201"
 /codon_start=1
 /protein_id="CAA50546.1"
 /db_xref="GI:1185072"
 /db_xref="SPTREMBL:Q64811"
 /translation="WHAFVIGKVDGTEPRYNSVLEHNRVAYVYVGGCFVCGALHI
 FSHLEFLGSRFVFKINCOIN"
 FSHLEFLGSRFVFKINCOIN

mRNA

mRNA

CDS

CDS

BASE COUNT 1943 a 1607 c 1399 g 2148 t
 ORIGIN

alignment_scores
 Quality: 87.50 Length: 252
 Ratio: 0.774 Gaps: 14
 Percent Similarity: 44.841 Percent Identity: 24.206

alignment_block
 US-09-151-189-2_COPY_441_676 x ACNPVDNA

Align seq 1/1 to: ACNPVDNA from: 1 to: 7297

16 SerCystGlyGlnLysThrLeuValHisArgPheAlaArgProGluAlaLe 32
 5512 TCTGGTGGAAGCGCATCGGTGGCGGTT 5540
 32 uGlyGlyThrLeuHisAsnThrIleAlaGlyAspLeuAspAlaAspPheA 49
 5541 TAACGGCGCGGACACAGTCACATACGCCGATCGGCGACACGCAAT 5590
 49 spLeuSerLeuLeuGluAsnAspGluLeuLeuArgValAlaGluLeu 65
 5591 ATTCAATGTTTGAATAACCG...TCACAACGTGATAACGTGCATC 5637

66 AsnAlaAlaGlnAsnProAsnGluValThrTyrLeu 78
 5638 ACCCGATCAGAACTCTGACACACAGTAGAGTGTCCGACGACTCCAG 5687
 79LeuproclualalelelnValGlySerProThrHisPro. 91
 5688 ATGCATAGATTACCCAC.....GGTACGGGCCACATGATAT 5725
 92SerTyrProSerGly 96
 5726 TCAACACGTTGACGACGATATTTCGTACACAGTGGCCAAATGGTGTGC 5775
 97HisAlaThrGlnAsnGlyAlaLaph 104
 5776 GATAATTTTGAAGTTATTTCCGACATCGAATGTGATCAATCAACGTGT 5825
 104 eAlaThrValLeuLysAlaLeulleGlyLeuAspArgGlyG 118
 5826 TGAACACGCGTGTATTATGGACAAATTTAGATTAAACATGCAA 5868
 118 LyGluCysPheProAsnProValPheProSerAspAspGlyLeuGluLeu 134
 5869TTCCCACTGAGTGTGACGGCAC 5895
 135 IleAsnPheGluGlyAlaCysLeuThrGlyGluGlyLeuLeuLeuLysLe 151
 5896CGGTGCGTCCACGCCAC 5912
 151 uAlaValAsaValAlaPheClyArgClnMetLeuGlyLe 164
 5913 CCGGACATGTCACCTTACCTTTAGCTTCCACGTTTGCCATTGAAATATC 5962
 165HisTyrArgPheAspGlyIleGlnGlyLeuLeuGlyGluThr 179
 5963 CAACACATATGTCATCGAC.....ATGCAAAACACCATTTGGCAGCAC 6009
 180 IleThrValArgThrLeuHisGlnGluLeuMetThrPheAlaGluGluAl 196
 6010 GAATGGTTTAAACAGTGTTCCTCAAGATTGTGTTTAAACACGACGC 6059
 196 aThrPheGluPheArgLeu 202
 6060 CATCTTTCCTCAATGGCTTTCTATCCGACAGACAAACACACATCGGC 6109
 203PheThrGlyGluValIleLysLeuPheGlnAspGlyThrPhe 216
 6110 TTACCCCTTCACCGCGGAGGTATCGACTGTTTGGAGACAACTTGATAC 6159
 217 SerIle 218
 6160 GATGTG 6165

seq_name: gb_p11:AB000134

seq_documentation_block:

LOCUS AB000134 909 bp DNA PLN 05-FEB-1999
 DEFINITION Prorocentrum micans mitochondrial DNA for cytochrome oxidase
 subunit I, partial cds.
 ACCESSION AB000134
 VERSION AB000134.1 GI:2897831
 KEYWORDS cytochrome oxidase subunit I
 SOURCE Prorocentrum micans (strain NIES12) mitochondrion DNA.
 ORGANISM Mitochondrion Prorocentrum micans
 Eukaryota; Alveolata; Dinophyceae; Prorocentrales; Prorocentraceae;

REFERENCE 1 (bases 1 to 909)

AUTHORS Inagaki,Y.

TITLE Direct Submission

JOURNAL

Submitted (27-DEC-1996) to the DDBJ/EMBL/GenBank databases. Yuji
 Inagaki, Biohistory Research Hall, Molecular Evolution Laboratory,
 Murasaki-cho 1-1, Takatsuki, Osaka 569, Japan
 (E-mail:yuji.inagaki@ims.brh.co.jp, Tel:0726-81-9761,

```

90 HisProSerTyrProSerGlyHisAlaThrGlnAsnGlyAlaPheAlaLeu 106
||||| 1438 CATCGTCTCTACGAGAGCGCCACGCTGTGCTGCGCAGTGTGAC 1487
106 rValLeuAlaLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 123
||||| 1488 GATCGTGAAGGCTTC.....TTCGACTCGGCATCGAGATC.....G 1525
123 snProValPheProSerAsp.....AspGlyLeuGluLeuLeuLeu 136
||||| 1526 ATCAGGTGTTCGAGGTGCAAGATGAGGACAGCTGTGTAAGTGTCT 1575
1526 rPheGluGlyAlaCysLeuThrTyrGluGlyLeuLeuLeuLeuLeu 153
||||| 1576 TTCAGGGAAGT...CTCACTGTGCGGTGAATGAACAGCTCGCGA 1622
153 lAsnValAlaPheGlyArgGlnMetLeuLeuLeuLeuLeuLeuLeu 170
||||| 1623 CAATATTCGATCGCGGTAAACATGCGCAGGTGTTCACTACTTCTCTGACC 1672
1623 lIleGlnGlyLeuLeuLeuGlyGluThrIleThrValArgThrLeuHis 186
||||| 1673 AGTTCGAGTCACTCTGCTCGGTGACGAGGTGCGATGGAATCTTGAA 1722
186 rGlnGluLeuMetThrPheAlaGluGluAlaThrPheGluPheArgLeu 203
||||| 1723 GAGCAAGTCTGACGTATGCGGAGAACTTCTTCTCACTTGCAGGAT 1772
203 eThrGlyGluValIleLeuLeu 210
||||| 1773 TGATGGAACTACATCCAGATC 1794

seq_name: N_Geneseq_36:V56021
seq_documentation_block:
ID V56021 standard; cDNA; 1791 BP.
AC V56021;
DE 09-DEC-1998 (first entry)
KW Haloperoxidase, CP.BP02 encoding cDNA.
KW Haloperoxidase; enzyme; CP.BP01; CP.BP02; vanadium; halide specificity;
OS Bromine; iodine; ds.
OS Corallina pilulifera.
FH Key Location/Qualifiers
FT CDS 1..1791
FT /tag= a
FT /product= "Haloperoxidase, CP.BP02"
FT /note= "the stop codon is not indicated"

J10248581-A.
22-SEP-1998.
06-MAR-1997: 070539.
06-MAR-1997: JP-070539.
(SAKA ) DTSUKA PHARM CO LTD.
WFL; 98-560733/48.
P-PSDB; W80531.
FT New haloperoxidase gene - and corresponding vector, transformed host
cell and method of preparation
PS Claim 2; Pages 15-18; 20pp; Japanese.
CC This cDNA encodes a haloperoxidase enzyme. The haloperoxidase genes
CC encoding the enzymes CP.BP01 and CP.BP02 are isolated from Corallina
CC pilulifera. A host cell transformed with a vector containing the
CC haloperoxidase genes can be used for the recombinant production of the
CC enzyme. The haloperoxidases of the invention are vanadium-dependent and
CC have halide specificity of bromine and iodine.
CC Sequence 1791 BP; 390 A; 459 C; 507 G; 435 T;
SQ

```

alignment_scores:

```

Quality: 318.00 Length: 231
Ratio: 2.224 Gaps: 8
Percent Similarity: 61.905

```

alignment_block:

```

us-09-151-189-2_copy_441_676 x V56021

```

```

Align seg 1/1 to: V56021 from: 1 to: 1791
1 HisTyrPheArgLeuLeu...GlyAlaAlaGluLeuAlaGlnArgAlaLeu 16
||||| 1132 CAGTCTGAGTCTGCTGACTGAGTGGCTACGCGCGCTTGAGGCGGT 1181
16 rCysTyrGlnLysTrpGlnValHisArgPheAlaArgProGluAlaLeu 33
||||| 1182 ACGTATCAGAAGTTTACATTCATGTCGCCGCGGCTGAGGCTACTG 1231
33 lGlyThrLeu.....HisAsnThrIleAlaGlyAspLeuAsp 45
||||| 1232 GTGGTCTGATTATCTTAACAAGAAATCTTCTCTCGGGTAGT..... 1275
45 AlaAspPheAspIleSerLeuLeuGluAsnAspGluLeuLeuLysArg 61
||||| 1276 .....GATATAATTCCTGAGTGTAGTGAAGTGGTGGAGAGCT 1316
61 ValAlaGluLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 72
||||| 1317 CTCATCATTCCTGGATGACGTTCTGAGAGCAATGAAAAACAGACAGGG 1366
72 snAsnGluVal.....ThrTyrLeuLeuProGluAlaLeu 83
||||| 1367 CTGACGGGATCTGACCGCGGATAAATCTTCTTCCGATGCGATT 1416
83 GlnValGlySerProThrHisProSerTyrProSerGlyHisAlaThrG 100
||||| 1417 GCGAGGGCAGCCCATTCATCTGCTTAAGTGTGAGTGGCAGCGTGGT 1466
100 nAsnGlyAlaPheAlaThrValLeuLysAlaLeuLeuLeu 113
||||| 1467 TGCTGCGCGATGTGTGACAAATCTCAAGGCGTTCTCGACGCCAACATTC 1516
114 LeuAspArgGlyGlyGluCysPheProAsnProValPheProSerAsp 129
||||| 1517 AGATCGATGAAGTGTTCGAG.....GTCGACACTGATGAG 1551
130 AspGlyLeuGluLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 146
||||| 1552 GACAACTTGTGAAGTCTGCTTTCAAGGGAAGT...CTCACTGTTCGCG 1598
146 lGluLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeuLeu 163
||||| 1599 TGAATTGAAACAGCTCCCGCAATCTTGGATCGGCGGAGACATGGCGG 1648
163 lIleHisTyrArgPheAspGlyIleGlnGlyLeuLeuLeuGlyGluThr 179
||||| 1649 GTGTCATCTATCTCTGACCACTTCGACTCTCTACTCTCTCTGAGACAG 1698
180 lIleThrValArgThrLeuHisGlnGluLeuMetThrPheAlaGluAl 196
||||| 1699 ATTGCGATGGAATCTTGGAGGAGCAGCGCTGACGTATGGCGAGAAGT 1748
196 aThrPheGluPheArgLeuPheThrGlyGluValIleLysLeu 210
||||| 1749 CTCTCTCAATTGCGAAATTCAGCGGAAGTACAAATCCAGATT 1791

seq_name: N_Geneseq_36:T03875
seq_documentation_block:
ID T03875 standard; cDNA; 2088 BP.
AC T03875;
DE 25-APR-1996 (first entry)
KW Chloroperoxidase cDNA.
KW Chloroperoxidase; haloperoxidase; antifouling paint; preservative;
OS Halide; chloride; ss.
OS Curvularia inaequalis.
FH Key Location/Qualifiers
FT CDS 7..1229
FT /tag= a
FT PN: W09527009-AL

```

12-OCT-1995.
31-MAR-1995; NL0123.
31-MAR-1994; EP-200893.
FR
24-JUN-1994; NL-001048.
PPA (SCHE-) STICHTING SCHEIKUNDIG ONDERZOEK IN NEDER.
PPA (TEWE-) STICHTING TECH.
PPA WETENSCHAPPEN.
PPA Dekker HL, Van Schijndel JWPM, Vollenbroek EGM, Wever R;
PPA DFTPI. WPI; 95-358602/45.
PPA DFTPI. P-PSDB; R82249.
PPA DFTPI. Use of halo-peroxidase enzymes - in antifoaming paints and methods
PPA DFTPI for determining the halide concn. of liquids
PPA DFTPI Example 3; Fig 6; 48pp; English.
PPA DFTPI A CDNA sequence (T03875) codes for a chloroperoxidase (R82249) of
PPA DFTPI Curvularia inaequalis. The cDNA was obt'd. by PCR using primers based
PPA DFTPI on known sequences of the gene and by 5'RACE. The apoprotein encoded
PPA DFTPI by the cDNA can be reactivated by addn. of vanadate. The cDNA can be
PPA DFTPI inserted into a vector for expression in prokaryotic host cells.
PPA DFTPI The recombinant chloroperoxidase is utilised in halide (esp.
PPA DFTPI chloride) analysis, or as an environmentally-friendly antifoaming
PPA DFTPI agent in ship paint.
PPA DFTPI Sequence 2088 BP. 465 A; 622 C; 531 G; 470 T;
PPA DFTPI

198 eGluPhear9.....LeuphethrGlycyluValleLysL 210
 1550 CGTGCTCCAGACCTAGAGACGATATAGGTACACACCCAGGGGGACGGTG 1699
 210 euPheSLAspGlyThrPheSerileAspGly 220
 1700 TGGACCCCGAGGGCGCTCTTCCCATCGGTGGT 1731
 seq_name: N_Geneseq_36.T13683
 seq_documentation_block:
 ID T13683 standard; DNA; 2520 BP.
 AC T13683;
 DE 05-SEP-1996 (first entry)
 DT ACNPV ORF 83, residues 67884-70427.
 KW Autographa californica nuclear polyhedrosis virus clone 6;
 KW disruption; non-essential gene; heterologous protein production;
 KW expression vector; baculovirus; ss
 KW Autographa californica nuclear polyhedrosis, virus clone 6.
 FH Key Location/Qualifiers
 FT 1..2520
 FT /'tag9' a
 FT /note= "corresponds to ACNPV nucleotides
 67884-70427"
 W09601320-A2.
 PD 18-JAN-1996.
 PF 30-JUN-1995, LB0578.
 PP 04-JUL-1994, GB-013420.
 PPA (NATU-) NATURAL ENVIRONMENT RES COUNCIL.
 PPI Ayres M, Bishop D, Possee R;
 PPI WPI: 96-087670/09.
 PPI GENBANK: L23858.
 DR Autographa californica nuclear polyhedrosis virus complete genome
 PT sequence - useful in the prodn. of vectors for enhanced
 PT heterologous protein expression, such as interleukin(s).
 PT interferon(s) and neurotoxin(s).
 PS Claim 1: Page 90-186; 122pp; English.
 PS T13636-731 show open reading frames 13, 20, 22-30, 32, 38, 41-46, 50-60,
 CC 62-63, 66, 68-79, 81-87, 91-92, 96-98, 101-103, 106-127, 129-130, 140-
 CC 146, 148-150, 152 and 154 from a total of 154 ORFs identified in the
 CC Autographa californica nuclear polyhedrosis virus (ACNPV) clone 6. Each
 CC gene is numbered according to its position in the virus genome beginning
 CC at the left end of the linear map, and irrespective of its orientation.
 CC The direction of transcription is relative to that of the polyhedrin
 CC gene. Of the 154 ORFs identified it was found that some of the ORFs (ORFs
 CC 27, 30, 32, 71, 86, 123, 126 and 137) are dispensable for virus
 CC replication in cell culture or insect larvae. These genes can be deleted
 CC from the genome to: (a) provide additional sites for inserting single or
 CC multiple copies of foreign genes; and (b) to reduce the size of the virus
 CC genome. The present sequence is designated ORF 83. Signal peptide and
 CC zinc finger motifs are present in the translated product.
 CC Sequence 2520 BP; 594 C; 527 G; 635 T.
 SQ

```

49 spileSerLeuLeuGluAsnAspGluLeuLeuLysArgValaLagluile 65
872 ATTTCAATGTTTGAATAATAACGAG...TCACAATGATAAGTGCATC 918
66 AsnAlaLagInAsnProAsnAsnGluValThrTyLeu..... 78
919 AACCGGATCAGAACTCTGCACACACGAGTACGAGTGTTCGCGGACATCCAG 968
79 .....LeuProGlnAlaLagLeuValGlySerProThrHisPro. 91
969 ATGCATAGATTATCCCAAC.....GGTACGGGCGCACATGAT 1006
92 .....SerTyProSerGly..... 96
1007 TCAACACGTTGACGACGATATTCGTACACACAGTGGCCAAATGGTGTGC 1056
97 .....HisAlaThrGlnAsnGlyAlaLaph 104
1057 GATAATTTTCAAGTATTTCGACATCGAATGTGATCAATCAACGTGT 1106
104 eAlaThrValLeu.....LysAlaLeuLeuLeuLeuLeuAspArgGly 118
1107 TGAACACGCGTGTGTATGACACAAATTTAGATTAAACATGCAA..... 1149
118 LysGluCysPheProAsnProValPheProSoraAspAspGlyLeuGluLeu 134
1150 .....TTCCCACTGAGGTGTT.....GACGCGACC..... 1176
135 IleAsnPheGluGlyAlaCysLeuThrTyGluGlyGluLeuLeuLysLeu 151
1177 .....GGTGTG.....GTGCGCCAC 1193
151 uAlaValAsnValAlaPheCysArgGlnMetLeuGlyLe..... 164
1194 CCGGACAAATGCACTTTTACGTCCACCGTTGCCATTTGAAATATTC 1243
165 .....HisTyraArgPheAspGlyIleGlnGlyLeuLeuGlyGluThr 179
1244 CAACACCAATATGCGATCGAC...ATGCAACCTCCATGTTGGCGACGAC 1290
180 IleThrValArgThrLeuHisGlnGluLeuMetThrPheAlaGluGluAl 196
1291 GAAATGGTTAAACAGTTGGTTCCAAAGATTGTGTTAAACACACGACG 1340
196 aThrPheGluPheArgLeu..... 202
1341 CATCTTTGCTCAATGGCTTTTGTATCGGAGACAGACAGACGCGCATCGGC 1390
203 .....PheThrGlyGluValIleLysLeuPheGlnAspGlyThrPhe 216
1391 TTAACCGGTTTCACGGCGGACCTTCGACTGTTTGGAGACAATTTGTAC 1440
217 SerIle 218
1441 GATGTG 1446

```

seq_name: N_Geneseq_36:TI3635

seq_documentation_block:

ID TI3635 standard; DNA: 133894 BP.

AC TI3635;

DT 03-SEP-1996 (first entry)

DE ACNPV genomic DNA clone 6.

KW Autographa californica nuclear polyhedrosis virus clone 6;

KW disruption; non-essential gene; heterologous protein production;

KW expression vector; baculovirus; ss

OS Autographa californica nuclear polyhedrosis virus clone 6.

PN W09601320-42.

PD 18-JAN-1996.

PF 30-JUN-1996; J05078.

PR 04-JUL-1994; GB-013420.

PA (NATU-) NATURAL ENVIRONMENT RES COUNCIL.

```

PI Ayres M, Bishop D, Possee R;
DR WPI; 96-087670/09.
PT GENBANK; L22858.
PT Autographa californica nuclear polyhedrosis virus complete genome
sequence - useful in the prodn. of vectors for enhanced
PT heterologous protein expression, such as interleukin(s),
PT interferon(s) and neurotoxin(s)
PS Disclosure: Page 90-186; 122pp: English.
CC The complete nucleotide sequence of the genome of clone 6 of the
CC baculovirus Autographa californica nuclear polyhedrosis virus (AcNPV)
CC has been determined. The sequence is taken from the Genbank record
CC L22858. The patent specification claims a polynucleotide selected from
CC open reading frames (ORFs) 13, 20, 22-26, 28-30, 32-35, 41-46, 50-60,
CC 62-63, 66, 68-79, 81-87, 91-92, 96-98, 101-103, 106-126, 129-130,
CC 140-146, 148-150, 152 and 154 from a total of 154 ORFs identified by
CC the patents. See TI3636-731. Expression vectors contg. the complete
CC genomic sequence of AcNPV, with the exception that at least one non-
CC essential ORF is disrupted or replaced are useful for the synthesis of
CC heterologous proteins.
SQ Sequence 133894 BP; 39195 A; 27151 C; 27347 G; 40201 T;

```

alignment_scores:

```

Quality: 86.50      Length: 252
Ratio: 0.765       Gaps: 14
Percent Similarity: 44.841 Percent Identity: 24.206

```

alignment_block:

US-09-151-189-2_COPY_441_676 x TI3635

align seg 1/1 to: TI3635 from: 1 to: 133894

```

16 SerCysTyrglnLysThrGlnValHisArgPheAlaArgProGluAlaLe 32
|||||.....
58700 TCGTGTGGGAA.....14.....CCGCATCGTCGCGTT 58728
32 uGlyGlyThrLeuHisAsnThrIleGlyAspLeuAspAlaAspPheA 49
|||||.....
58729 TAACGGCGCGGACACACGATACATACGCCCATATGGCGACACGCAAT 58778
49 spileSerLeuLeuGluAspGlyLeuLeuLysArgValaLagluile 65
|||||.....
58779 ATTCAATGTTTGAATAATAACGAG...TCACAATGATAAGTGCATC 68825
66 AsnAlaLagInAsnProAsnAsnGluValThrTyLeu..... 78
|||||.....
58826 AACCGGATCAGAACTCTGCACACGAGTGTTCGCGGACATCCAG 58875
79 .....LeuProGlnAlaLagLeuValGlySerProThrHisPro. 91
|||||.....
58876 ATGCATAGATTATCCCAAC.....GGTACGGGCGCACATGAT 58913
92 .....SerTyProSerGly..... 96
|||||.....
58914 TCACACGTTGACGACGATATTCGTACACACAGTGGCCAAATGGTGTGC 58963
97 .....HisAlaThrGlnAsnGlyAlaLaph 104
|||||.....
58964 GATAATTTTGAAGTATTTCGACATCGAATGTGATCAATCAACGTGT 59013
104 eAlaThrValLeu.....LysAlaLeuLeuLeuLeuLeuAspArgGly 118
|||||.....
59014 TGAACACGCGTGTGTATGACACAAATTTAGATTAAACATGCAA..... 59056
118 LysGluCysPheProAsnProValPheProSoraAspAspGlyLeuGluLeu 134
|||||.....
59057 .....TTCCCACTGAGGTGTT.....GACGCGACC..... 59083
135 IleAsnPheGluGlyAlaCysLeuThrTyGluGlyGluLeuLysLeu 151
|||||.....
59084 .....GGTGTG.....GTGCGCCAC 59100
151 uAlaValAsnValAlaPheGlyArgGlnMetLeuGlyLe..... 164

```



```

69101 CGCGGACAAATGCAACATTTTACGTTCCAGCTTTCATGCAATGAATAATTC 69150
165 .....NiaTyrArgPheAspGlyTleGlnGlyLeuLeuLeuGlyThrPhe 179
165151 CAAACGATTATGGCATGCAC...ATGCACACCTCCATGTGGGACGACGC 69197
180 IleThrValArgThrLeuHisGlnGluLeuMetThrPheAlaGluGlnAl 196
69198 GAAATGTTAAACAGCTGTGTTCCAGAGATTGTCGTTAAACACGACGC 69247
196 aThrPheGluPheArgLeu..... 202
69248 CATCTTTCCTCATGGCTTTTGTATCGGAGACAAAGACGCCATCGGGC 69297
203 .....PheThrGlyGluValIleLysLeuPheLysPheLysPheLysPhe 216
69298 TTAACCGGTTCACGGGAGCGCTATCGACTGTTTGGAGACAACTTGTAC 69347
217 SerIle 218
69348 GATGTG 69353

```

seq_name: N_Geneseq_36:T00803

seq_documentation_block:

```

ID T00803 standard; DNA; 1830 BP.
AC T00803;
DE 27-APR-1996 (first entry)
KW Chloroperoxidase gene.
KW antimicrobial; antiseptic; ss.
OS Curvularia inaequalis strain CBS102.42.
PN W9527046-A.
PD 12-OCT-1995.
PR 31-MAR-1995; E01229.
PA (UNIL) UNILEVER NV.
PA (UNIL) UNILEVER PLC.
PI Barnett P, Hondmann DH, Simona LH, Ter Steeg PF,
PI Wever R,
DR WPI: 95-358625/46.
PT Enzymatic antimicrobial compsn, esp. disinfectant - contg. vanadium
PT halo:peroxidase, esp. from Curvularia inaequalis, halide and
PT hydrogen peroxide source
PS Claim 11; Fig 2: 48pp; English.
CC A gene (T00803) encoding the vanadium chloroperoxidase of Curvularia
CC inaequalis CBS102.42 was isolated from a genomic library using a
CC probe obtd. by amplification of cDNA using primers based on isolated
CC fragments of the gene. The gene can be inserted into a vector and
CC used to produce recombinant chloroperoxidase in host cells, e.g.
CC Saccharomyces cerevisiae. The enzyme is used together with a
CC halide source and H2O2 in antimicrobial compositions active against
CC e.g. Streptococcus faecalis and food spoilage microorganisms.
SQ Sequence 1830 BP; 402 A; 584 C; 461 G; 383 T;

```

alignment_scores:

```

Quality: 85.50 Length: 138
Ratio: 1.379 Gaps: 5
Percent similarity: 44.928 Percent identity: 21.739

```

alignment_block:

US-09-151-189-2_COPY_441_676 x T00803

Align aeg 1/1 to: T00803 from: 1 to: 1830

```

65 IlesnAlaAlaGlnAsnProhAsnGluValThrTyrLeuLeuProG1 81
1141 CTCGGTCCCGCCACTACTAACACCAACGACATCCATTC..... 1179
81 nAlaIleGlnValGlySerProThrHisProSerTyrProSerGlyHisA 98

```

```

1180 .....AGCCCTCCTTTCCAGCTTACCCATCGTGGTCAGC 1213
98 IaThrGlnAsnGlyAlaPheAlaThrValLeuLys..... 109
1214 CGACCTTTGGCGGTCTGTGTTTCCAAATGGTGGCTCGATACATAACAGCG 1263
109 ..... 109
1264 CGCGTAGGTACATGGAAGCAGCAGACCCGACCAACATTGCCATCGATAT 1313
110 .....AlaLeuIleGlyLeuAspArgGlyGlyGlyLucyPheP 122
1314 CATGATCTCGAGGAGCTCAACGGCGTGAACGGGACCTACGCGACGCTT 1363
122 roAsnProValPheProSerAspArgGlyLeuGluLeuLeu..... 135
1364 ATGACCCCGACGGCCCAATCGAAGACCAACCGGATATCGGCGACCCGC 1413
136 .....AsnPheGluGlyAlaLys.....LeuThrTyrGluGlyGluL 148
1414 ATTGTTCGCGACTTCGACTCGGCTGGGAACCTCATGTTCGAAGACCCAT 1463
148 eAsnLysLeuAlaValAsnValAlaPheGlyArgGlnMetLeuGlyIleH 165
1464 TTCGCGCATC.....TTCCTCGGTGTGTC 1486
165 IsTyrArgPheAsp 169
1487 ACTGGCGTTTCGAT 1500

```

seq_name: N_Geneseq_36:T51612

seq_documentation_block:

```

ID T51612 standard; DNA; 2822 BP.
AC T51612;
DE 30-APR-1997 (first entry)
DE Curvularia verruculosa halo:peroxidase gene.
KW Halo:peroxidase; halide oxidation; halogenation; antimicrobial;
KW disinfectant; ss.
OS Curvularia verruculosa strain CBS 147.63.
PN Key
PN Location/Qualifiers
FT cds
FT primer_bld
FT complement (477..496)
FT /tag= b;
FT /note= primer aHapI binding site
FT primer_bld
FT /tag= c
FT /note= primer aHapI binding site*
FT WO9704102-A1.
FT 06-FEB-1997.
FT 09-JUL-1996; U11458.
FT 14-JUL-1995; US-001194.
FT 21-FEB-1996; US-603534.
FT (NOVO) NOVO NORDISK BIORECH INC.
FT (NOVO) NOVO-NORDISK AS.
FT Berka RW, Cherry J, Fuglsang C, Halkier T, Oxenboll KM,
FT WPI: 97-132641/12.
FT P-PSDB; W12042.
FT Halo:peroxidase from Curvularia verruculosa - useful for oxidn. of
FT halide, halogenation and, in presence of hydrogen peroxide and
FT thiocyanate, as antimicrobial
FT Claim 28; Page 33-35; 58pp; English.
CC The halo:peroxidase gene (T51612) of Curvularia verruculosa CBS.63
CC codes for an enzyme (W12042) that shows optimum activity at about
CC 60 deg C and pH 5.5, which retains at least 75% activity after
CC incubation for 1 hr at pH 7.0 and 60 deg C in the presence of 0.1%
CC H2O2, and which prefers bromide over chloride as substrate. It was
CC isolated from a genomic library using a PCR amplified partial clone
CC (see also T51613-14) as probe. The 2822 bp fragment in a positive
CC plaque has been deposited (in PHAP4.1, in E. coli DH10B) as NRRL
CC B-21519. The gene can be used to express the halo:peroxidase in
CC transformed host cells for use in halide oxidation, cpo.

```

CC halogenation and (with H2O2 and thiocyanate) as an antimicrobial.

50 Sequence 2823 BP; 651 A; 808 C; 674 G; 689 T;

alignment_scores:

Quality: 83.50 Length: 138
Ratio: 1.347 Gaps: 5
Percent Similarity: 44.928 Percent Identity: 21.014

alignment_block:

US-09-151-189-2_COPY_441_676 x T51612

Align seg 1/1 to: T51612 from: 1 to: 2822

65 ILeaAlaAlaGlnAsnProAsnGlnValThrThrLeuLeuProG 81

1608 CTCGGTCCCGCAGCTACAAACACAAACGACATACCTTC..... 1646

81 hAlaAlaGlnValGlySerProThrH1sProSerTyProSerGlyH1sA 98

1647AGCCTCCTTTCCCGCCCTACCCCACTCGGCCAGC 1680

98 laThrGlnAsnGlyAlaPheAlaThrValLeuLys..... 109

1681 CCACCTTGGCGGTGCTGATTCCAGATGGTCCCGCCGCTACTAGAACGGG 1730

109 109

1731 CCGCTAGGCACCTGGAGGACGACGACGACGACGACGACGACGACGACGAC 1780

110AlaLeuIleGlyLeuAspArgGlyGlyGlyGlyGlyGlyGlyGly 122

1781 GATGATATCCGAGGAGCTCAACGGCGTGAACCGGCACTCGCCGACCCCT 1830

122 TGAAsnProValPheProSerAspGlyLeuGluLeuLys..... 135

1831 ACGACCGGAGTCCCGCCGCTCGAGACCAACGAGGATGCTCGCACCCGC 1880

136AsnPheGluGlyAlaCys...LeuThrTyGluGlyGlu 148

1881 ATCGTGGCCCATTTGCTACGCTGGGGAATGATGTCGAAACGCCAT 1930

148 eAsnLysLeuAlaValAsnValAlaPheGlyArgGlnMetLeuGlyIleH 165

1931 TCTCGGATC.....TTCTCGGCTCC 1953

165 IsTyArgPheAsp 169

1954 ACTGCCGCTTCGAT 1967

seq_name: N_Geneseq_36.V21209_11

seq_documentation_block:

Continuation (12 of 17) of V21209 from base 1100001 (Methanococcus jannaschii circular

WP Sequence split into 17 fragments LOCUS V21209 Accession V21209

Fragment Name Begin End

WP V21209_00 1 110000

WP V21209_01 100001 210000

WP V21209_02 200001 310000

WP V21209_03 300001 410000

WP V21209_04 400001 510000

WP V21209_05 500001 610000

WP V21209_06 600001 710000

WP V21209_07 700001 810000

WP V21209_08 800001 910000

WP V21209_09 900001 1010000

WP V21209_10 1000001 1110000

WP V21209_11 1100001 1210000

WP V21209_12 1200001 1310000

WP V21209_13 1300001 1410000

WP V21209_14 1400001 1510000

WP V21209_15 1500001 1610000

WP V21209_16 1600001 1664976

alignment_scores:

Quality: 82.50 Length: 91
Ratio: 2.012 Gaps: 2
Percent Similarity: 45.055 Percent Identity: 25.275

alignment_block:

US-09-151-189-2_COPY_441_676 x V21209_11

Align seg 1/1 to: V21209_11 from: 1 to: 110000

91 ProSerTyProSerGlyHisAlaThrGlnAsnGlyAlaPheAlaThrVa 107

33250 CCACGCTTCCCACTGGTCATACAACTTTAGCATTACATAGCAACATC 33299

107 lLeuLysAlaLeuIleGlyLeuAspArgGlyGlyGlyGlyGlyGlyGlyGly 124

33300 CTTA..... 33303

124 roValPheProSerAspGlyLeuGluLeuIleAsnPheGluGlyAla 140

33303..... 33303

141 CysLeuThrTyGluGlyGluIleAsnLysLeu.....AlaVa 153

33304 ..TTATTTACTCAAAACAACTTGAATATGTTTAACTTGGGCTAT 33350

153 lAsnValAlaPheGlyArgGlnMetLeuGlyIleHisTyArgPheAspG 170

33351 NATTTAGCTTATAGTACGTTTATGTCGACTTCATTATCCTTGGATC 33400

170 lylleGlnGlyLeuLeuGly 177

33401 TCCTTGCTGAATGATTATGGA 33423

seq_name: N_Geneseq_36.T45009

seq_documentation_block:

ID T45009 standard; DNA; 735 BP.

AC T45009:

DT 13-AUG-1997 (first entry)

DE Serratia marcescens IAM 13540 acid phosphatase DNA.

KW TAM 13540; acid phosphatase; production; nucleoside; 5'-phosphate;

KN ester; condiment; pharmaceutical; intermediate; ds.

OS Serratia marcescens.

FH Key location/Qualifiers

FT cds 1..735

FT /tag- a

FT /product- acid_phosphatase

FN W09637603-A1.

PD 28-NOV-1996.

PF 24-MAY-1996; J01402.

PR 25-MAY-1995; JP-149781.

PR 26-MAR-1996; JP-094680.

PA (AJIN) AJINOMOTO CO INC.

PI Asano Y, Mihara Y, Utagawa T, Yamada H;

DR WP1; 97-021215/02.

DR P-PSDB; W06460.

PT Efficient production of nucleoside 5'-phosphate - by reaction of a

PT nucleoside with a phosphoric acid donor in the presence of an acid

PT phosphatase

PS Example 24: Pages 74-75; 95pp; Japanese.

CC The present sequence encodes the Serratia marcescens IAM 13540

CC acid phosphatase (AP), which can be used to produce a nucleoside

CC 5'-phosphate ester from the corresponding nucleoside when a

CC phosphate donor, e.g. poly, phenyl or carbamyl phosphoric acid,

CC is reacted in its presence at pH 3.0 to 5.5. The PA can be used for

CC the economic and efficient production of nucleoside 5'-phosphate for

CC esters for use as condiments, pharmaceuticals and intermediates for

CC pharmaceuticals.

SC Sequence 735 BP; 180 A; 215 C; 198 G; 142 T;

alignment_scores: Quality: 81.50 Length: 210
Ratio: 0.823 Gaps: 10
Percent Similarity: 47.143 Percent Identity: 23.333

alignment_block:
US-09-151-189-2_copy_441_676 x T45009
Align seg 1/1 to: T45009 from: 1 to: 735

13 GlnArgAlaSerCysTyrGlnLysTrpGlnValBlaArgPheAlaArgPr 29
198 GAAATAGTGGCAATACCTCGGCTGCAAGCAGGCTTATGATGACGCC 247
29 oGluAlaLeuGly GlyThrLeuHlaAnThrIlaAlaGlyAspLeuasp 45
248 AGGTGCGGGGAGCGCGTTCGCCGCGCATTTTCC 282
46 AlaAspPhe.....AspIleSerLeuLeuGluAsnAspGluLeuLeu 60
283 AACGCTTCGGCTAGAAATAGCCCAACGGAAGCGCGAGCTGTAA 332
60 sArgValAlaGluLeuAsnAlaAlaGlnAsp 71
333 GCTGTGTGTAATGCTGCAAGCAGCGCGGATTTGCGGACCGCAGCG 382
71 roAsnAsnGluValThrTyrLeuProGlnAlaIleGlnValGlySer 87
383 CCAAAATACATATATGCGCATTCGCCCTTTCGCTTTTATACGAAGCG 432

88 ProThrHisPro.....SertYrPr 94
433 ACCTGCGGACCGGAGAAAGCACCCTGTCGAAGACGGTCTTACCC 482
94 oSerGlyHisAlaThrGlnAspGlyAlaPheAlaThrValLeuLysAla 111
483 TTCGCGCATACCACTGCGCTGCGGCGCCGCTGCTGCTGCTGCTGCTG 532
111 euileGlyLeuAspArgGlyGlyLeuGlyCysPheProAsnProValPhePro 127
533 TC.....AACCC.....540
128 SerAspAspGlyLeuGluLeuLeuAsnPheGluGlyAlaCysLeuThr 144
541GCCAG 545
144 tGluGlyGluLeuAsnLysLeuAlaValAsnValAlaPheGlyArgGlu 161
546 CCAGGTGAAATCTGCGCGGCTATGATGCGCAAGCGGGTTA 595
161 tLeuGlyIleHisTyrArgPheAspGlyIleGlnGlyLeuLeuLeu 177
596 TCTGCGGTATCATGCGCAAGCGAC...GTGACTGCGCGCGCGCTGCGG 642
178 GluThrIleThrValArgThrLeuHisGlnGlnLeuMetThrPheAla 193
643 CGGTGCGCGCATGCTGCGGCTGTCATGCCGAA...CCACCTTCGCCCC 689
194GluAlaLeuThrPheGluPhe 200
690 CCAGTGCAGAAAGCGCAAGCAATTC 717

seq_name: N_Geneseq_36:V43062

seq_documentation_block:
ID V43062 standard; DNA; 735 BP.
AC V43062;
DT 21-OCT-1998 (first entry)
DE DNA encoding an acid phosphatase enzyme.
KW acid phosphatase; preparation; nucleoside 5'-phosphate ester;
KW seasoning intermediate; ds.
OS Serratia ficaria.

Key Location/Qualifiers
CDS 1..735
/*tag= a

EP-857788-A2.
12-AUG-1998. 309365.
20-NOV-1997. JP-161674.
18-JUN-1997. JP-161674.
21-NOV-1996. JP-311103.
(AJIN) AJINOMOTO CO INC.
Asano Y, Mihara Y, Utagawa T, Yamada H;
WPI: 98-416010/36.
P-PSDB: W71033.
Preparation of nucleoside 5'-phosphates comprises reacting nucleoside with phosphate donor in presence of acid phosphatase - used as seasonings or pharmaceutical intermediates
Example 24; Pages 47-48; 83pp; English.
The present sequence encodes an acid phosphatase enzyme. The specification describes a method for the preparation of nucleoside 5'-phosphate esters. The method comprises reacting a nucleoside with a phosphate donor at pH 3.0-5.5 in the presence of an acid phosphatase that has been altered to increase its affinity for the nucleoside and/or to increase its thermal stability, or in the presence of a microorganism that has been transformed with recombinant DNA containing a gene coding for such an acid phosphatase. Nucleoside 5'-phosphates are useful as seasonings or pharmaceuticals or as intermediates for them.
CC note: this sequence appears to be claimed (claim 6), but as the claim refers to amino acid sequences, it is clear that the corresponding protein is being claimed.
CC Protein is being claimed.
SQ Sequence 735 BP; 180 A; 215 C; 198 G; 142 T;

alignment_scores:
Quality: 81.50 Length: 210
Ratio: 0.823 Gaps: 10
Percent Similarity: 47.143 Percent Identity: 23.333

alignment_block:
US-09-151-189-2_COPY_441_676 x V43062

Align seg 1/1 to: V43062 from: 1 to: 735

13 GlnArgAlaSerCysTyrGlnLysTrpGlnValHisArgPheAlaArgPr 29
198 GAAATAGTGGCAATACCTCGGCTGCAAGCAGGCTTATGATGACGCC 247
29 oGluAlaLeuGly GlyThrLeuHlaAnThrIlaAlaGlyAspLeuasp 45
248 AGGTGCGGGGAGCGCGTTCGCCGCGCATTTTCC 282
46 AlaAspPhe.....AspIleSerLeuLeuGluAsnAspGluLeuLeu 60
283 AACGCTTCGGCTAGAAATAGCCCAACGGAAGCGCGAGCTGTAA 332
60 sArgValAlaGluLeuAsnAlaAlaGlnAsp 71
333 GCTGTGTGTAATGCTGCAAGCAGCGCGGATTTGCGGACCGCAGCG 382
71 roAsnAsnGluValThrTyrLeuProGlnAlaIleGlnValGlySer 87
383 CCAAAATACATATATGCGCATTCGCCCTTTCGCTTTTATACGAAGCG 432
88 ProThrHisPro.....SertYrPr 94
433 ACCTGCGGACCGGAGAAAGCACCCTGTCGAAGACGGTCTTACCC 482
94 oSerGlyHisAlaThrGlnAsnGlyAlaPheAlaThrValLeuLysAla 111
483 TTCGCGCATACCACTGCGCTGCGGCGCCGCTGCTGCTGCTGCTGCTG 532
111 euileGlyLeuAspArgGlyGlyLeuGlyCysPheProAsnProValPhePro 127
533 TC.....AACCC.....540

Percent Similarity: 47.783 Percent Identity: 20.690

alignment_block:

US-09-151-189-2_COPY_441_676 x V52179

Align seg 1/1 to: V52179 from: 1 to: 7577

```

37 HisAsnThrIleAlaGlyAspLeuAspAlaAspPheAspIleSerLeuLeu 53
   ::::::::::::::::::::
4943 TACACTGCCATTCATAGAACTTTGACCTCCCTTATGATATCATCTGCC 4992
   ::::::::::::::::::::
53 uGluAsnAsp.....GluLeuLeuLysArgValAlaG 64
   ::::::::::::::::::::
4993 CGAATCTGAAACCCAGGACACAGATTGAGATTTTACAAAATAAGTGG 5042
   ::::::::::::::::::::
64 luLeuAsnAlaAlaGlnAsnProAsnAsnGluValThrTyLeuLeuPro 80
   ::::::::::::::::::::
5043 AACTG.....GCCGAATCTCTAGCCTCTCTCGT 5071
   ::::::::::::::::::::
81 GlnAlaIleGlnValGlySerProThrHisProSer.....Ty 93
   ::::::::::::::::::::
5072 ATCGTCCCAATCCCTTGTGTCAGTCCGACGAGGATATAATCGCTTTTATTT 5121
   ::::::::::::::::::::
93 rProSerGlyHisAlaThrGlnAsnGlyAlaPheAlaThrValLeuLysA 110
   ::::::::::::::::::::
5122 TCCAAAGAAATGTTGGACCAAAACAGCATCTTTCGAAGCATTTACCCAGC 5171
   ::::::::::::::::::::
110 laLeuIleGlyLeuAspArgGlyGlyGlyCysPheProAsnPro..... 124
   ::::::::::::::::::::
5172 AATTTTACACTTGATTGAGACGATCATTTTACCCCAATAAATAAGTAG 5221
   ::::::::::::::::::::
125 .....ValPheProSerAspAspGlyLeuGluLeuIleAsnPheGlu 139
   ::::::::::::::::::::
5222 AAGAGAAATTTAGTATTA.....CtRATTCGGTTGATTCGACGG 5268
   ::::::::::::::::::::
139 yAlaCysLeuThrTyrgluGlyGluIleAsnLys..... 150
   ::::::::::::::::::::
5269 AACCCTTTCACAGCCAAAGGAAATCATCTCTGAGTTTTCCTCCCA 5318
   ::::::::::::::::::::
151 .....LeuAlaValAsnValAlaPheGlyArg 159
   ::::::::::::::::::::
5319 TCCAGATGCCAAAGAGAGCTGGTCAAAAGTCGTGATTCGAACCTGCCCG 5368
   ::::::::::::::::::::
160 GlnMetLeuGlyIleHisTyraGpPheAspGlyIleGln..... 172
   ::::::::::::::::::::
5369 CCTATCGCAGCGCTTGCCAAACTCTAGACGACTTCAGATTGACAGACGA 5418
   ::::::::::::::::::::
173 .....GlyLeuLeuGlyGluThrIleT 181
   ::::::::::::::::::::
5419 GGGGACTATGTGGTAACCTTCAACGGTGCCTTGTCCAGAAACTGCTA 5468
   ::::::::::::::::::::
181 hrValArgThrLeuHisIleGluLeuMetThrPheAlaGluAlaThr 197
   ::::::::::::::::::::
5469 CAGGACATGAGATTATCAGCGATCTTGACTATATGAGGATTATCTAGAT 5518
   ::::::::::::::::::::
198 PheGluPhe 200
   ::::::::::::::::::::
5519 ATGGGAATTC 5527

```

seq_name: N_Geneseq_36:X18867

seq_documentation_block:

ID X18867 standard: DNA; 12019 BP.

AC X18867;

DT 10-MAY-1999 (first entry)

DE Alcaligenes sp. Pox proteins R, A, B, C, D, E, F, G, H and I coding DNA.

KW Alcaligenes; PoxR; PoxA; PoxB; PoxC; PoxD; PoxE; PoxF; PoxG; PoxH;

KX PoxI; aromatic; oxidative; petroleum purification; chemical industry;

KY drug; ds.

OS Alcaligenes sp.

PN J11042088-A.

PD 16-FEB-1999.

PF 25-JUL-1997; 200625.

PR 25-JUL-1997; JP-200625.

PA (TOFU) TONEN CORP.

DR WPI; 99-197820/17.

DR P-PDOB; W98969, W98970, W98971, W98972, W98973, W98974, W98975,

DR W98976, W98977, W98978.

PT New encoding an aromatic cpd. oxidative decompn. enzyme - useful in

PT the fields of petroleum purification, chemical and drug industries.

PS Claim 1: Page 7-19; 35pp; Japanese.

CC The present sequence encodes Alcaligenes sp. proteins PoxR, PoxA, PoxB,

CC PoxC, PoxD, PoxE, PoxF, PoxG, PoxH and PoxI. The proteins are useful in

CC the fields of petroleum purification, chemical industry and drug

CC industry related to the synthesis, conversion and decomposition of

CC aromatic compounds.

SO Sequence 12019 BP; 2015 A; 4104 C; 4151 G; 1747 T;

alignment_scores:

Quality: 80.50 Length: 226

Ratio: 0.694 Gaps: 15

Percent Similarity: 51.327 Percent Identity: 23.009

alignment_block:

US-09-151-189-2_COPY_441_676 x X18867

Align seg 1/1 to: X18867 from: 1 to: 12019

```

1 HistyPheATGLeuIleGlyAlaAlaGluLeuAlaGlnATGAlaSer... 16
   ::::::::::::::::::::
9672 CACTATATCGAGCTGCTCGGCTCCAGGAGACGCGCGGACGCGCAGGG 9721
   ::::::::::::::::::::
17 .....CysTyrglnLys.....TrpGlnValHisArgPheAlaArgProGluA 31
   ::::::::::::::::::::
9722 CCGCGTCTACTGAGGCGCTGGACGACGACGACCAT..... 9758
   ::::::::::::::::::::
31 laLeuGlyGlyThrLeuHisAsnThrIleAlaGlyAspLeuAspAla... 46
   ::::::::::::::::::::
9759 .....CNCAGCGTGTCTGCGCGGCGCCACGACGCGCG 9791
   ::::::::::::::::::::
47 .....AspPheAspIleSerLeuLeuGluAsnAspGluLeuLeuLysAr 51
   ::::::::::::::::::::
9792 GGCAATGCATCATCATGGCTTCGCGTCGACAGCGATGCCACGCTGGAGCG 9841
   ::::::::::::::::::::
61 gValAlaGluIleAsnAlaAlaGlnAsnProAsnAsnGluValThrTyrl 78
   ::::::::::::::::::::
9842 GCTGGCGCGCGAAGTCCCGCCAGCGGCTGGCGCAGGACTCGCCTCGA 9891
   ::::::::::::::::::::
78 eu...LeuProGlnAlaIleGlnValGlySerProThrHisProSerTy 93
   ::::::::::::::::::::
9892 TCCCGCGCGCGACGACCTCGCGACCGCGCGGCGCTTCGCTTCACCAT 9941
   ::::::::::::::::::::
94 ProSerGlyHisAlaThrGlnAsnGlyAlaPheAlaThrValLeuLysAl 110
   ::::::::::::::::::::
9942 CCACGCGCGCATGCCATCGAG.....CTGTCGCGGAC.....AAGA 9979
   ::::::::::::::::::::
110 aLeuIleGlyLeuAspArgGlyGlyGlyCysPheProAsnProValPhe 127
   ::::::::::::::::::::
9980 CAAGCTCGGCTGC...CAGACCGCGCGACCTGAATCGGACCCC...TGGC 10023
   ::::::::::::::::::::
127 roSerAsp...AspGlyLeuLeuLeuLeuAsnPheGluGlyAlaCysLeu 142
   ::::::::::::::::::::
10024 CGGACGACCTCGCGCGCATGCCCGGACGCGCTTCGAC...CACTGCGCTG 10070
   ::::::::::::::::::::
143 ThrTyrgluGlyLeuIleAsnLysLeuAlaValAsnValAlaPheGlyAr 159
   ::::::::::::::::::::
10071 CTGTACGCGCGACCTCAACGCG.....ACGGTCAAGCTCTTCCG 10111
   ::::::::::::::::::::
159 gGlnMetLeuGlyIleHisTyraGpPheAspGlyIleGlnGly..... 173
   ::::::::::::::::::::
10112 TGACGTGCTGCGCTCTCGCTCGCGCGGAGAGTGTGGCGGCGCGGAGC 10161
   ::::::::::::::::::::
174 ..LeuLeuLeuGlyGluThrIleThrValArgThrLeuHisGlnGluLeu 189
   ::::::::::::::::::::
10162 GGATGGTGTATGCCCGCCCTTCCTGACCTGCGAACCAAGCGCGCACCATC 10211

```


OM of: US-09-151-189-2_COPY_441_676 to: Issued_Patents_NA.* out_format : pfs
 Date: May 17, 2000 10:57 AM
 About: Results were produced by the GenCore software, version 4.5,
 Copyright (c) 1993-2000 CompuGen Ltd.

Command line parameters:
 -MODEL=fract p2n model -DEV=rlp
 -O=/cgn2_1/USPTO/US09151189/runat_15052000_082513_1631/app_query.fasta.1
 -DB=Issued_Patents_NA -FASTA=FASTA -SUFFIX=441-676.rni
 -GAPOP=12.000 -GAPEXT=4.000 -MINMATCH=0.100 -LOOPCL=0.000
 -LOOPEXT=0.000 -CGAPOP=4.000 -CGAPEXT=0.050 -XGAPOP=10.000
 -XGAPEXT=0.500 -FGAPOP=6.000 -FGAPEXT=7.000 -YGAPOP=10.000
 -YGAPEXT=0.500 -DELOP=6.000 -DELEXT=7.000 -START=1
 -MATRIX=DIOSUM62 -TRANS=human40.cdi -LIST=45 -DOCALIGN=200
 -THR_SCORE=pct -ALIGN=15 -MODE=LOCAL -OUTFMT=pfs -NORM=ext
 -MINLEN=0 -MAXLEN=1000000 -USER=US09151189 -NCPU=6 -ICPU=3
 -NO_XLPXY -WAIT -THREADS=1

Search information block:
 Query: US-09-151-189-2_COPY_441_676
 Query length: 236
 Database: Issued_Patents_NA.*
 Database sequences: 226296
 Database length: 63486255
 Search time (sec): 41.480000

score_list:	Strd Orig	ZScore	EScore Len	Documentation
/cgn2_6/ptodata/1/lna/5C.COMB.seq:US-08-679-405-1 +	83.50	146.22	1.10	2822
/cgn2_6/ptodata/1/lna/5D.COMB.seq:US-08-842-799-1 +	83.50	146.22	1.10	2822
/cgn2_6/ptodata/1/lna/5B.COMB.seq:US-08-842-799-1 +	83.50	146.22	1.10	2822
/cgn2_6/ptodata/1/lna/5A.COMB.seq:US-08-842-799-1 +	83.50	146.22	1.10	2822
/cgn2_6/ptodata/1/lna/5C.COMB.seq:US-08-750-145A-23 +	81.50	157.82	0.2486	735
/cgn2_6/ptodata/1/lna/5C.COMB.seq:US-08-750-145A-23 +	81.50	157.82	0.2486	735
/cgn2_6/ptodata/1/lna/5C.COMB.seq:US-08-624-601-3 +	75.50	138.48	2.97	1188
/cgn2_6/ptodata/1/lna/5C.COMB.seq:US-07-931-943-1 +	75.50	127.74	11.79	2912
/cgn2_6/ptodata/1/lna/5C.COMB.seq:US-08-624-601-1 +	75.50	127.74	11.79	2912
/cgn2_6/ptodata/1/lna/5B.COMB.seq:US-08-070-1659-5 +	75.00	127.62	11.96	2675
/cgn2_6/ptodata/1/lna/5B.COMB.seq:US-08-070-1659-5 +	75.00	127.62	11.96	2675
/cgn2_6/ptodata/1/lna/5D.COMB.seq:US-08-070-1659-9 +	75.00	127.59	12.00	2681
/cgn2_6/ptodata/1/lna/5D.COMB.seq:US-08-070-1659-9 +	75.00	127.59	12.00	2681
/cgn2_6/ptodata/1/lna/5C.COMB.seq:US-08-619-198-2 +	74.00	120.03	31.65	4170
/cgn2_6/ptodata/1/lna/5C.COMB.seq:US-08-619-198-2 +	74.00	120.03	31.65	4170
/cgn2_6/ptodata/1/lna/5C.COMB.seq:US-08-399-561-1 +	74.00	118.03	40.91	4928
/cgn2_6/ptodata/1/lna/5C.COMB.seq:US-08-619-198-8 +	74.00	117.39	44.33	5192
/cgn2_6/ptodata/1/lna/5C.COMB.seq:US-08-619-198-1 +	74.00	117.39	44.33	5192
/cgn2_6/ptodata/1/lna/5D.COMB.seq:US-08-387-942C-1 +	73.50	105.66	200.07	12588
/cgn2_6/ptodata/1/lna/5A.COMB.seq:US-08-330-492-2 +	73.50	127.00	12.95	1931
/cgn2_6/ptodata/1/lna/5B.COMB.seq:US-08-687-379-5 +	71.50	118.51	38.50	2955
/cgn2_6/ptodata/1/lna/5B.COMB.seq:US-08-687-379-5 +	71.50	118.51	38.50	2955
/cgn2_6/ptodata/1/lna/5C.COMB.seq:US-08-790-912-1 +	71.50	109.62	120.37	6201
/cgn2_6/ptodata/1/lna/5A.COMB.seq:US-07-731-157A-1 +	71.00	121.11	27.55	2163
/cgn2_6/ptodata/1/lna/5A.COMB.seq:US-08-541-780-1 +	71.00	121.11	27.55	2163
/cgn2_6/ptodata/1/lna/5D.COMB.seq:US-08-481-206-3 +	70.50	114.20	66.92	3505
/cgn2_6/ptodata/1/lna/5D.COMB.seq:US-08-481-206-3 +	70.50	114.20	66.92	3505
/cgn2_6/ptodata/1/lna/5B.COMB.seq:US-08-486-269A-3 +	70.50	129.36	9.57	819
/cgn2_6/ptodata/1/lna/5B.COMB.seq:US-08-221-750A-8 +	69.50	121.07	27.70	1635
/cgn2_6/ptodata/1/lna/5B.COMB.seq:US-08-387-156-9 +	69.50	121.07	27.70	1635
/cgn2_6/ptodata/1/lna/5B.COMB.seq:US-08-694-865-9 +	69.50	121.07	27.70	1635
/cgn2_6/ptodata/1/lna/5D.COMB.seq:US-08-878-748-9 +	69.50	121.07	27.70	1635
/cgn2_6/ptodata/1/lna/5C.COMB.seq:US-09-124-491-9 +	69.50	119.86	32.36	1809
/cgn2_6/ptodata/1/lna/6.COMB.seq:US-09-124-491-9 +	69.50	119.86	32.36	1809
/cgn2_6/ptodata/1/lna/5C.COMB.seq:US-08-694-865-15 +	69.50	118.06	40.76	2102
/cgn2_6/ptodata/1/lna/5C.COMB.seq:US-09-124-491-15 +	69.50	118.06	40.76	2102
/cgn2_6/ptodata/1/lna/5B.COMB.seq:US-08-455-559-9 +	69.50	116.83	47.72	2329
/cgn2_6/ptodata/1/lna/5C.COMB.seq:US-08-455-559-9 +	69.50	116.83	47.72	2329
/cgn2_6/ptodata/1/lna/5A.COMB.seq:US-07-960-932-1 +	69.50	114.65	63.13	2794
/cgn2_6/ptodata/1/lna/5A.COMB.seq:US-07-908-253-1 +	69.50	114.65	63.13	2794
/cgn2_6/ptodata/1/lna/5B.COMB.seq:US-08-455-970A-1 +	69.50	114.65	63.13	2794
/cgn2_6/ptodata/1/lna/5B.COMB.seq:US-08-387-156-5 +	69.50	114.65	63.13	2794

/cgn2_6/ptodata/1/lna/5C.COMB.seq:US-08-694-865-5 + 69.50 114.65 63.13
 /cgn2_6/ptodata/1/lna/5D.COMB.seq:US-08-878-748-5 + 69.50 114.65 63.13
 /cgn2_6/ptodata/1/lna/5D.COMB.seq:US-08-535-837-1 + 69.50 114.65 63.13
 /cgn2_6/ptodata/1/lna/6.COMB.seq:US-09-124-491-5 + 69.50 114.65 63.13
 2794

seq_name: /cgn2_6/ptodata/1/lna/5C.COMB.seq:US-08-679-405-1
 seq_documentation_block:
 : Sequence 1: Application US/08679405
 : Patent No. 5866393
 : GENERAL INFORMATION:
 : APPLICANT: Fugisang, Claus
 : APPLICANT: Halkier, Torben
 : APPLICANT: Oxenboll, Karen M.
 : APPLICANT: Berka, Randy M.
 : APPLICANT: Cherry, Joel
 : TITLE OF INVENTION: Haloperoxidases from Curvularia
 : NUMBER OF SEQUENCES: 21
 : CORRESPONDENCE ADDRESS:
 : ADDRESSEE: No. 58663930 No. 58663930th America, Inc.
 : STREET: 405 Lexington Avenue, Suite 6400
 : CITY: New York
 : STATE: New York
 : COUNTRY: U.S.A.
 : ZIP: 10174
 : COMPUTER READABLE FORM:
 : MEDIUM TYPE: Floppy disk
 : COMPUTER: IBM PC compatible
 : OPERATING SYSTEM: PC-DOS/MS-DOS
 : SOFTWARE: Patent in Release #1.0, Version #1.30
 : CURRENT APPLICATION DATA:
 : APPLICATION NUMBER: US/08/679,405
 : FILING DATE: July 9, 1996
 : CLASSIFICATION: 435
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: 60/001,194
 : FILING DATE: July 14, 1995
 : PRIOR APPLICATION DATA:
 : APPLICATION NUMBER: 08/603,534
 : FILING DATE: February 21, 1996
 : ATTORNEY/AGENT INFORMATION:
 : NAME: Lambiris, Elias J.
 : REGISTRATION NUMBER: 33,728
 : REFERENCE/DOCKET NUMBER: 4441.210.US
 : TELECOMMUNICATION INFORMATION:
 : TELEPHONE: (212) 867-0123
 : TELEFAX: (212) 878-9655
 : INFORMATION FOR SEQ ID NO: 1:
 : SEQUENCE CHARACTERISTICS:
 : LENGTH: 2822 base pairs
 : TYPE: nucleic acid
 : STRANDEDNESS: single
 : TOPOLOGY: linear
 : MOLECULE TYPE: DNA (genomic)
 : PEPTIDE:
 : NAME/KEY: CDS
 : LOCATION: 477..2276
 : US-08-679-405-1

alignment_scores:
 : Quality: 83.50 Length: 138
 : Ratio: 1.947 Gaps: 5
 : Percent Similarity: 44.928 Percent Identity: 21.014

alignment_block:
 : US-09-151-189-2_COPY_441_676 x US-08-679-405-1
 : Align seg 1/1 to: US-08-679-405-1 from: 1 to: 2822
 : 65 llesnAlaAlaGlnProAsnGluValThrTyLeuProL 81
 :|||.....
 :|||.....

```
1608 CTGGTGGCCCGCTACAAACAGACATACCTTC.....1646
81 alalleGlnValGlySerProThrHisProSerTyProSerGlyHisA 98
   |||||
1647 .....AAGCCTCTTCCCGCGCTACCATCTGCCACG 1680
98 lethrGlnAsnGlyAlaPheAlaThrValLeuLys.....109
   |||||
1681 CCACCTTGGCGTCTCTATTCCAGATGCTCGCGCTACTACACGGG 1730
109 .....109
1731 CGGTAGGACCTGGAAGCAGCAGACACAGACATGCCATGCAT 1780
110 .....1780
1781 GATCATATCCGAGGAGCTCAACGGCGTGAACCGGCTGCGCCAGCCCT 1830
122 roasnProValPheProSerAspGlyLeuGluLeulle.....135
   |||||
1831 ACAGCCGAGCTCCCGCTACGAGACCAACAGCAGTATCGTCCGACCGC 1880
136 .....136
1881 ATCGTGGCGCCTTGTACTAGCTGCGTGGAAATGATGTTGAAACGCCAT 1930
148 easnLysLeuAlaValAsnValAlaPheGlyArgGlnMetLeuGlyLeuH 165
   |||||
1931 TTCTCGCATC.....TTCTCGCGCTCC 1953
165 lstyrArgPheAsp 169
1954 ACTGGCGCTTCGAT 1967

seq_name: /cgn2_6/ptodata/1/ina/5D_COMB.seq:US-08-842-799-1

seq_documentation_block:
; Sequence 1, Application US/08842799
; Patent No. 5965418
; GENERAL INFORMATION:
; APPLICANT: Fuglsang, Claus
; APPLICANT: Halkier, Torben
; APPLICANT: Oxenboll, Karen M.
; APPLICANT: Berka, Randy M.
; APPLICANT: Cherry, Joel
; TITLE OF INVENTION: Haloperoxidases from Curvularia
; TITLE OF INVENTION: Verruculosa and Nucleic Acids Encoding same
; NUMBER OF SEQUENCES: 21
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 59654180 No. 5965418disk of No. 5965418th America, Inc.
; STREET: 405 Lexington Avenue, Suite 6400
; CITY: New York
; STATE: New York
; COUNTRY: U.S.A.
; ZIP: 10178
; COMPUTER READABLE FORM: disk
; MEDIUM TYPE: Floppy
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/842.799
; FILING DATE: 16-APR-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/679,405
; FILING DATE: July 9, 1996
; APPLICATION NUMBER: 60/001,194
; FILING DATE: July 14, 1995
; PRIOR APPLICATION DATA: 08/603,534
; APPLICATION NUMBER: 08/603,534
; FILING DATE: February 21, 1996
; ATTORNEY/AGENT INFORMATION:
```

```
NAME: Lambiris, Elias J.
REGISTRATION NUMBER: 33,728
REFERENCE/DOCKET NUMBER: 4441.210.US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 867-0123
TELEFAX: (212) 878-9655
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2822 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLSCULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 477..2276
US-08-842-799-1

alignment_scores:
Quality: 83.50 Length: 138
Ratio: 1.347 Gaps: 5
Percent Similarity: 44.928 %Percent Identity: 21.014

alignment_block:
US-09-151-189-2_copy_441_676 x US-08-842-799-1
Align seg 1/1 to: US-08-842-799-1 from: 1 to: 2822
65 llaeAlaAlaGlnAsnProAsnGluValThrTyLeuLeuProGl 81
   |||||
1608 CTGGTGGCCCGCTACAAACAGACATACCTTC.....1646
81 naAlleGlnValGlySerProThrHisProSerTyProSerGlyHisA 98
   |||||
1647 .....AAGCCTCTTCCCGCGCTACCATCTGCCACG 1680
98 lathrGlnAsnGlyAlaPheAlaThrValLeuLys.....109
   |||||
1681 CCACCTTGGCGTCTCTATTCCAGATGCTCGCGCTACTACACGGG 1730
109 .....109
1731 CGGTAGGACCTGGAAGCAGCAGACACACATGCCATGCAT 1780
110 .....1780
1781 GATCATATCCGAGGAGCTCAACGGCGTGAACCGGCTGCGCCAGCCCT 1830
122 roasnProValPheProSerAspGlyLeuGluLeulle.....135
   |||||
1831 ACAGCCGAGCTCCCGCTACGAGACCAACAGCAGTATCGTCCGACCGC 1880
136 .....136
1881 ATCGTGGCGCCTTGTACTAGCTGCGTGGAAATGATGTTGAAACGCCAT 1930
148 easnLysLeuAlaValAsnValAlaPheGlyArgGlnMetLeuGlyLeuH 165
   |||||
1931 TTCTCGCATC.....TTCTCGCGCTCC 1953
165 lstyrArgPheAsp 169
1954 ACTGGCGCTTCGAT 1967

seq_name: /cgn2_6/ptodata/1/ina/PTCUS_COMB.seq:PCT-US96-11458-1

seq_documentation_block:
; Sequence 1, Application PCT/TUS9611458
; GENERAL INFORMATION:
; APPLICANT:
; APPLICANT:
; TITLE OF INVENTION: Haloperoxidases from Curvularia
```

TITLE OF INVENTION: Verruculosa and Nucleic Acids Encoding Same

NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESSEE: NOVO Nordisk of North America, Inc.
STREET: 405 Lexington Avenue, Suite 6400
CITY: New York
STATE: New York
COUNTRY: U.S.A.
ZIP: 10174

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentid Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US96/11458
FILING DATE: 9-JUL-1996
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/001,194
FILING DATE: 14-JUL-1995
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/603,534
FILING DATE: 21-FEB-1996
ATTORNEY/AGENT INFORMATION:
NAME: Lambiris, Elias J.
REGISTRATION NUMBER: 33,728
REFERENCE/DOCKET NUMBER: 4441.204-WO
TELEPHONE: (212) 867-0123
TELEFAX: (212) 878-9655
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 2822 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 477..2276
PCT-US96-11458-1

alignment_scores:
Quality: 83.50 Length: 138
Ratio: 1.347 Gaps: 5
Percent similarity: 44.928 Percent identity: 21.014

alignment_block:

US-09-151-189-2_COPY_441_676 x PCT-US96-11458-1

Align seq 1/1 to: PCT-US96-11458-1 from: 1 to: 2822

65 llaasAlaAlaGlnAspProThrHisProSerGlyHisA 81
1608 CTGGGCCCCAGCTACAAACAAACGACATACCTTC 1646
81 nAlaLeGlnValGlySerProThrHisProSerGlyHisA 98
1647AAGCTCTCTTCCCGCGCTACCATCTGCCACG 1680
98 lathGlnAsnGlyAlaPheAlaThrValLeuLys 109
1681 CCACCTTGGCGTCTCTATTCCAGATGTCGCGCTACTACACGG 1730
109 109
1731 CGCGTAGGCACTGGAGGACGACGACACACATTCCTGAT 1780
110AlaLeuLeGlyLeuAspArgGlyGlyLucysPheP 122
1781 GATGATATCGAGGAGCTCAACGCGGTGAACCGCACCTGCGCGCCCT 1830

122 toAspProValPheProSerAspArgGlyLeuGluLeuLeu 135
1831 ACGACCGGAGTGGCCCGCATCGAAGACCAACCAAGTATCGTCCGACCCGC 1880
136AspPheGluGlyAlaCys...LeuThrTyrGluGlyGlu 148
1881 ATCTGGCGCCACCTTCACCTACGCTGGGAATGATGTTCGAAAACCCAT 1930
148 eAsnLysLeuAlaValAsnValAlaPheGlyArgGlnMetLeuGlyLys 165
1931 TTCGCGATCTTCTCGCGGCTCC 1953
165 lTyrArgPheAsp 169
1954 ACTGGCGCTTCGAT 1967

seq_name: /cgn2_6/ptodata/1/ina/6_COMB.seq:US-08-750-145A-23

seq_documentation_block:

Sequence 23, Application US/08750145A

Patent No. 6010851

GENERAL INFORMATION:

APPLICANT: MIHARA, Yasuhiko

APPLICANT: UTAGAWA, Takashi

APPLICANT: YAMADA, Hideaki

APPLICANT: ASANO, Yasuhisa

TITLE OF INVENTION: Method for producing Nucleoside-5'

TITLE OF INVENTION: Phosphate Ester

NUMBER OF SEQUENCES: 24

CORRESPONDENCE ADDRESS:

ADDRESS: OSLON, SEIVAK, MCLELLAND, MAIER & NEUSTADT, P.C.

STREET: 1755 S. JEFF. DAVIS HIGHWAY, SUITE 400

CITY: ARLINGTON

STATE: VA

COUNTRY: USA

ZIP: 22202

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentid Release #1.0, Version #1.30

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/750,145A

FILING DATE: 01-JAN-1997

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP 7-149781

FILING DATE: 05-May-1995

PRIOR APPLICATION DATA:

APPLICATION NUMBER: JP 8-094680

FILING DATE: 26-Mar-1996

ATTORNEY/AGENT INFORMATION:

NAME: NORMAN F. OBLON

REGISTRATION NUMBER: 24,618

REFERENCE/DOCKET NUMBER: 0010-0830-0 PCT

TELECOMMUNICATION INFORMATION:

TELEPHONE: 703-413-3000

TELEFAX: 703-413-2220

INFORMATION FOR SEQ ID NO: 23:

SEQUENCE CHARACTERISTICS:

LENGTH: 735 base pairs

TYPE: nucleic acid

STRANDEDNESS: double

TOPOLOGY: linear

MOLECULE TYPE: DNA (genomic)

HYPOTHETICAL: NO

ANTI-SENSE: NO

ORIGINAL SOURCE:

ORGANISM: Serratia ficaria

STRAIN: IAM 13540

FEATURE:

NAME/KEY: CDS

LOCATION: 1..732
US-08-750-145A-23

alignment_scores:
Quality: 81.50 Length: 210
Ratio: 0.823 Gaps: 10
Percent Similarity: 47.143 Percent Identity: 23.333

alignment_block:
US-09-151-189-2_COPY_441_676 x US-08-750-145A-23

Align seg 1/1 to: US-08-750-145A-23 from: 1 to: 735

```
13 GlnArgAlaSerCysTyrGlnLysTrpGlnValHisArgPheAlaArgPr 29
198 GAAATAGTGCCTAGTACTCCGCTGCACAGAGGCTTATGTCGCCCC 247
29 OGluAlaLeuGly.GlyThrLeuHisThrIleAlaGlyAspLeuAsp 45
248 ACCTTCGCGGAGCGC.....GTTCGCGCGCATTTTC 282
46 AlaAspPhe.....AspIleSerLeuGluAsnAspGluLeuLeu 60
283 AACGCTTCGCGCTAGAAATAGTACCCACAGGAAACCGCGAGCTGTAA 332
60 sArgValAlaGluIleAsn.....AlaAlaGlnAsn 71
333 GCTGTGATGAAATGCGTGAAGAGCGCGCATTTGCGCGACCGCAGCG 382
71 roAsnAsnGluValThrTyrLeuLeuProGlnAlaIleGlnValGlySer 87
383 CCAATACATATGCGCATTCGCGCTTCGCTTTATTAACGAGCG 432
88 ProThrHisPro.....SerTyrPr 94
433 ACTGCGCGGAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 482
94 OSerGlyHisAlaThrGlnAsnGlyAlaPheAlaThrValLeuLysAla 111
483 TTCGCGCATACACACATCGCTGGCGAGCGCGCTGGCTGGCTGAA 532
111 eulleGlyLeuAspArgGlyGlyGlyCysPheProAsnProValPhePro 127
533 TC.....AACCC 540
128 SerAspAspGlyLeuGluIleAsnPheGluGlyAlaCysLeuThrTy 144
541 .....GCCAG 545
144 xGluGlyGluIleAsnLysLeuAlaValAsnValAlaPheGlyArgGln 161
546 GCAGGTGAATCTCGAGCGCGCTATGATGCGCAACCGCGGTA 595
161 etLeuGlyIleHisTyrArgPheAspGlyIleGlnGlyLeuLeuGly 177
596 TCTGCGGTATACCTGGCAAGAGCGAC...GTGACTGCGCGCGCATGCG 642
178 GluThrIleThrValArgThrLeuHisGlnGluLeuMetThrPheAla 193
643 CGCTGCGCATGCTGCGCGCTTTGCGATGCCGAA...CCACCTTCGCGCG 689
194 .....GluGluAlaThrPheGluPhe 200
690 CCACCTCAAAAGGCCAAGACGATTC 717
```

seq_name: /cgn2_6/ptodata/1/lna/6_COMB.seq:US-08-750-145A-23

seq_documentation_block:
Sequence 27, Application US/08975698A
Patent No. 6015697
GENERAL INFORMATION:
APPLICANT: MIHARA, YASUHIRO

APPLICANT: UTAGAWA, TAKASHI
APPLICANT: YAMADA, HIDEAKI
TITLE OF INVENTION: METHOD FOR PRODUCING NUCLEOSIDE-5'-PHOSPHATE
TITLE OF INVENTION: METHOD FOR PRODUCING NUCLEOSIDE-5'-PHOSPHATE
NUMBER OF SEQUENCES: 30
CORRESPONDENCE ADDRESS:
ADDRESS: OBLON, SPIVAK, MCCLELLAND, MAIER & NEUSTADT, P.C.
STREET: 1755 S. JEFFERSON DAVIS HIGHWAY, FOURTH FLOOR.
CITY: ARLINGTON
STATE: VIRGINIA
COUNTRY: USA
ZIP: 22202
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/975,698A
FILING DATE: 21-NOV-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: OBLON, NORMAN F
REGISTRATION NUMBER: 24,618
REFERENCE/DOCKET NUMBER: 0010-0895-0
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703) 413-3000
TELEFAX: (703) 413-2220
INFORMATION FOR SEQ ID NO: 27:
SEQUENCE CHARACTERISTICS:
LENGTH: 735 base pairs
TYPE: nucleic acid
STRAINEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: genomic DNA
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE: Serratia ficaria
ORGANISM: IAM 13540
STRAIN: IAM 13540
FEATURE:
NAME/KEY: CDS
LOCATION: 1..732
US-08-975-698A-27

alignment_scores:
Quality: 81.50 Length: 210
Ratio: 0.823 Gaps: 10
Percent Similarity: 47.143 Percent Identity: 23.333

alignment_block:
US-09-151-189-2_COPY_441_676 x US-08-975-698A-27

Align seg 1/1 to: US-08-975-698A-27 from: 1 to: 735

```
13 GlnArgAlaSerCysTyrGlnLysTrpGlnValHisArgPheAlaArgPr 29
198 GAAATAGTGCCTAGTACTCCGCTGCACAGAGGCTTATGTCGCCCC 247
29 OGluAlaLeuGly.GlyThrLeuHisThrIleAlaGlyAspLeuAsp 45
248 ACCTTCGCGGAGCGC.....GTTCGCGCGCATTTTC 282
46 AlaAspPhe.....AspIleSerLeuGluAsnAspGluLeuLeu 60
283 AACGCTTCGCGCTAGAAATAGTACCCACAGGAAACCGCGAGCTGTAA 332
60 sArgValAlaGluIleAsn.....AlaAlaGlnAsn 71
333 GCTGTGATGAAATGCGTGAAGAGCGCGCATTTGCGCGACCGCAGCG 382
```


181 ThrValArgThrLeuHisGln.....Glule 189
1003 ACCGAGCGCGCTTTCATCAGTTCAAGGCTCGCGCGGTCCTTCGTT 1052
189 uKetThrPheAlaGluGlu.....AlaThrPheGluPheArgL 202
1053 TTGCTCTATGTCGAGGATTGGTTACACCTCTGTTTGACTTCCTCC 1102
202 euphenr 204
1103 TCTTTCT 1110

seq_name: /cgn2_6/ptodata/1/1na/5A_COMB.seq:US-07-931-943-1
seq_documentation_block:
; Sequence 1, Application US/07931943
; Patent No. 5470729
; GENERAL INFORMATION:
; APPLICANT: KAPER, James B.
; APPLICANT: BAUDRY-MAURELLI, Bernadette
; APPLICANT: FASANO, Alessio
; TITLE OF INVENTION: METHOD OF ISOLATING RESTRICTION FRAGMENT
; TITLE OF INVENTION: DELETIONS IN VIBRIO CHOLERAE, AND PRODUCTS THEREOF
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: BAKER & MCKENZIE
; STREET: 815 Connecticut Avenue, N.W.
; CITY: Washington
; STATE: DC
; COUNTRY: US
; ZIP: 20006
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patent Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/931,943
; FILING DATE: 19920812
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/821,072
; FILING DATE: 16-JAN-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/533,315
; FILING DATE: 05-JUN-1990
; APPLICATION NUMBER: US 06/581,406
; FILING DATE: 17-FEB-1984
; APPLICATION NUMBER: US 06/472,276
; FILING DATE: 04-MAR-1993
; APPLICATION NUMBER: US 07/363,383
; FILING DATE: 05-JUN-1989
; APPLICATION NUMBER: US 06/867,633
; FILING DATE: 27-MAY-1986
; ATTORNEY/AGENT INFORMATION:
; NAME: Kile, Bradford E.
; REGISTRATION NUMBER: 25,223
; REFERENCE/DOCKET NUMBER: BANC20016P6
; TELEPHONE: (202) 452-7000
; TELEFAX: (202) 452-7074
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 2912 base pairs
; TYPE: NUCLEIC ACID
; STRANDEDNESS: double
; TOPOLOGY: unknown
; MOLECULE TYPE: DNA (genomic)
; ORIGINAL SOURCE:

ORGANISM: Vibrio cholerae
STRAIN: CLASSICAL 395
FEATURE:
NAME/KEY: CDS
LOCATION: 1034..2218
FEATURE:
NAME/KEY: CDS
LOCATION: 2221..2508
US-07-931-943-1

alignment_scores:
Quality: 75.50 Length: 236
Ratio: 0.668 Gaps: 14
Percent Similarity: 47.881 Percent Identity: 23.729

alignment_block:
US-09-151-189-2_COPY_441_676 x US-07-931-943-1
Align seg 1/1 to: US-07-931-943-1 from: 1 to: 2912
32 LeuGlyGlyThrLeuHisAsnThrIleAlaGlyAspLeuAspAlaAspPh 48
1478 ATGCTTCCACTCTC...AATCATGTCAGTGGTCAGGTGCCACTCTCA 1524
48 eAspIleSerLeuLeuGluAsn.....AspGluLeuLeuLysA 61
1525 AGATATGGTTCAGCTCATACGAGTTCACGGCGGCGGTGTTCTGAAA 1574
61 tGValAlaGluIleAsnAlaAlaGlnAsnProAsnAsnGluValThrTy 77
1575 AGTGAACGCAATCAC.....ATCGATTGATGGCAGATACATAT 1618
78 LeuLeuProGlnAlaIle..... 83
1619 CTTGAAGAGTTTCGATCGATGATGATGGATACACAGCGAGGTCAAAA 1668
84GlnValGlySerProThrHisProSerT 93
1669 AGCCAGGATGAACCTCTCTTCACGTGTGTGTGTGTGTGTGTGTGT 1718
93 yProSerGlyHisAlaThrGlnAsnGlyAlaPheAlaThrValLeuLys 109
1719 ACCAACTAAGATGCT.....GTGCTTCGG 1744
110 AlaLeuIleGlyLeu.....AspArgGlyGlyGlyCysPhe..... 121
1745 CGATGATGAGCTTAAGATTCACCTCGTGGGTGTCTGCTCCGCTAA 1794
122ProAsnProValPheProSerAspAspGlyLeuGluLeu 135
1795 CCCAGACCAACCAATCCACGCCGCGATAGCAGCAGC.....C 1835
135 LeAsnPheGluGlyAlaCysLeuThrTyGlyGlyGlyLeuAsnLysLeu 151
1836 CCAATTATACAGGGCGCTTAATACCATCTCTAAAGAGCTCATACCTTA 1885
152AlaValAsnValAlaPhe.....GlyAr 159
1886 GAGACGATTCACAGCACTCGACACCATGACACGCGCTATCAGGGG 1935
159 GlnMetLeuGlyIleHisTyArgPhe..... 168
1936 CTGTAGTAACTCTCGCTGCTGCTGCTGCTGCTGCTGCTGCTGCTG 1985
169AspGlyIleGlnGlyLeu.....LeuLeuGlyGluThrIle 180
1986 AGTAGAAGCGCTCAGCAGATTTAAAGCAGATGATCAACGATAAATC 2035
181 ThrValArgThrLeuHisGln.....Glule 189
2036 ACCAGTCGGCTTTGTCATCAGTTCAAGGCTCGCGCGGTCGCTTCGT 2085

```

189 uMetThrPheAlaGluGlu.....AlaThrPheGluPheArgL 202
      : : : : : : : : : : : : : : : : : : : : : : : :
2086 TTGCTCCATGTCAGGAGTTGTTACAACTCTGTTGACTTCCTCC 2135
      : : : : : : : : : : : : : : : : : : : : : : : :
202 euPheThr 204
      : : : : : : : : : : : : : : : : : : : : : : : :
2136 TCTTTCT 2143
      : : : : : : : : : : : : : : : : : : : : : : : :

```

seq_name: /cgn2_6/ptodata/1/ina/5C_COMB.seq:US-08-624-601-1

```

seq_documentation_block:
: Sequence 1, Application US/08624601
: Patent No. 5892653
: GENERAL INFORMATION:
: APPLICANT: Kaper Dr., James B.
: TITLE OF INVENTION: Vibrio cholerae O1 (CV111) and non-O1
: TITLE OF INVENTION: (CV112 and CV112RM) serogroup vaccine strains, methods
: TITLE OF INVENTION: of making same and products thereof
: NUMBER OF SEQUENCES: 16
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Spencer & Frank
: STREET: 1100 New York Ave. N.W. Suite 300 East
: CITY: Washington
: STATE: D.C.
: COUNTRY: USA
: ZIP: 20005
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC Compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentia Release #1.0, version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/624.601
: FILING DATE: 08-APR-1996
: CLASSIFICATION: 424
: ATTORNEY/AGENT INFORMATION:
: NAME: Scheller Dr., John W.
: REGISTRATION NUMBER: 26,031
: REFERENCE/DOCKET NUMBER: BANCZ0019P2
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (202)414-4000
: TELEFAX: (202)414-4040
: INFORMATION FOR SEQ ID NO: 1:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2912 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: DNA (genomic)
: HYPOTHETICAL: NO
: ANTI-SENSE: NO
: ORIGINAL SOURCE:
: ORGANISM: Vibrio cholerae
: STRAIN: classical 395
: IMMEDIATE SOURCE:
: CLONE: ace/orfu
: US-08-624-601-1

```

alignment_scores:
 Quality: 75.50 Length: 236
 Ratio: 0.668 Gaps: 14
 Percent Similarity: 47.881 Percent Identity: 23.729

alignment_block:

US-09-151-189-2_COPY_441_676 x US-08-624-601-1

Align seg 1/1 to: US-08-624-601-1 from: 1 to: 2912

```

32 LeuGlyThrLeuHisAsnThrIleAlaGlyAspLeuAspAlaAsp 48
      : : : : : : : : : : : : : : : : : : : : : : : :
1478 ATGGCTCGACTCTC...AATCATGTCAGTGTGAGTGTCCCACTCTCA 1524
      : : : : : : : : : : : : : : : : : : : : : : : :

```

```

48 eAspSerLeuLeuGluAsn.....AspGluLeuLeuLysA 61
      : : : : : : : : : : : : : : : : : : : : : : : :
1525 AGATATGGTTCAGCTCAATAGAGTTTCACGCGGACCGTGTCTTGA 1574
      : : : : : : : : : : : : : : : : : : : : : : : :
61 rGValAlaGluIleAsnAlaAlaGlnAsnProAsnAsnGluValThr 77
      : : : : : : : : : : : : : : : : : : : : : : : :
1575 AAGTGACGCAATCAAC.....AATCGATTGATGGCAGATAAATAT 1618
      : : : : : : : : : : : : : : : : : : : : : : : :
78 LeuLeuProGlnAlaIle..... 83
      : : : : : : : : : : : : : : : : : : : : : : : :
1619 CTGTAGAGAGTTCGATCGATCGATGATGATGATGATGATGATGAT 1668
      : : : : : : : : : : : : : : : : : : : : : : : :
84 .....GlnValGlySerProThrHisProSerT 93
      : : : : : : : : : : : : : : : : : : : : : : : :
1669 AGCCAAAGATGACTCTCTTCACGCTGTGTGTTCTGTCACACGATTT 1718
      : : : : : : : : : : : : : : : : : : : : : : : :
93 YProSerGlyValAlaThrGlnAsnGlyAlaPheAlaThrValLeuLys 109
      : : : : : : : : : : : : : : : : : : : : : : : :
1719 ACCAAGTAAGATGCT.....GTCTTCGG 1744
      : : : : : : : : : : : : : : : : : : : : : : : :
110 AlaLeuIleGlyLeu.....AspArgGlyGlyGluCysPhe..... 121
      : : : : : : : : : : : : : : : : : : : : : : : :
1745 GCGATTGATGACTTAAGATTACCTCGCTGGGTGTTGCTTCGCGCTAA 1794
      : : : : : : : : : : : : : : : : : : : : : : : :
122 .....ProAsnProValPheProSerAspAspGlyLeuGluLeuI 135
      : : : : : : : : : : : : : : : : : : : : : : : :
1795 CCAGACCAACCAATCCACGCGCACCCGATAGCAGCAGC.....C 1835
      : : : : : : : : : : : : : : : : : : : : : : : :
135 LeAsnPheGluGlyAlaCysLeuThrTyxGluGlyGluIleAsnLys 151
      : : : : : : : : : : : : : : : : : : : : : : : :
1836 CCAATATACAGGGGCGCTTAATACCATCTCTAATAAGCTCAATACCTTA 1885
      : : : : : : : : : : : : : : : : : : : : : : : :
152 .....AlaValAsnValAlaPhe.....GlyAr 159
      : : : : : : : : : : : : : : : : : : : : : : : :
1886 GAGAGGATTCACAGCAACTCGACACCATGAACACGCGGCTATCAGG 1935
      : : : : : : : : : : : : : : : : : : : : : : : :
159 gGlnMetLeuGlyIleHisTyxArgPhe..... 168
      : : : : : : : : : : : : : : : : : : : : : : : :
1936 CTGTAGTAACTCTCTGCTGTGCTGCTGCTGCTGCTGCTGCTGCTG 1985
      : : : : : : : : : : : : : : : : : : : : : : : :
169 .....AspGlyIleGlnGlyLeu.....LeuLeuGlyGluThrIle 180
      : : : : : : : : : : : : : : : : : : : : : : : :
1986 AGTTAGAAACGGCTCAGCAGAAATTAAGCAGATGATCAACGATAAATC 2035
      : : : : : : : : : : : : : : : : : : : : : : : :
181 ThrValArgThrLeuHisGln.....GluLe 189
      : : : : : : : : : : : : : : : : : : : : : : : :
2036 ACCAGTCGGCTTTCATCAGTTCAAGGCTCGGCGGCGTGCCTTCGTT 2085
      : : : : : : : : : : : : : : : : : : : : : : : :
189 uMetThrPheAlaGluGlu.....AlaThrPheGluPheArgL 202
      : : : : : : : : : : : : : : : : : : : : : : : :
2086 TTGCTCCATGTCAGGAGTTGTTACAACTCTGTTGACTTCCTCC 2135
      : : : : : : : : : : : : : : : : : : : : : : : :
202 euPheThr 204
      : : : : : : : : : : : : : : : : : : : : : : : :
2136 TCTTTCT 2143
      : : : : : : : : : : : : : : : : : : : : : : : :

```

seq_name: /cgn2_6/ptodata/1/ina/5B_COMB.seq:US-08-070-165F-5

```

seq_documentation_block:
: Sequence 5, Application US/08070165F
: Patent No. 5750365
: GENERAL INFORMATION:
: APPLICANT: Chiu, Ing-Ming
: APPLICANT: Poulin, Matthew L
: TITLE OF INVENTION: Acidic Fibroblast Growth Factor (aFGF)
: NUMBER OF SEQUENCES: 12
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Ing-Ming Chiu
: STREET: 52052 Davis Medical Research Center, 480 West
: STREET: 9th Avenue
: CITY: Columbus
: STATE: Ohio

```



```

: TELEFAX: (614)-293-5631
: INFORMATION FOR SEQ ID NO: 5:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2675 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: HYPOTHETICAL: NO
: ANTI-SENSE: NO
: ORIGINAL SOURCE:
: ORGANISM: No. 592528ophthalmus viridescens
: DEVELOPMENTAL STAGE: Adult
: TISSUE TYPE: Regenerating forelimb blastema
: CELL TYPE: Mesenchyme and Epithelium
: IMMEDIATE SOURCE:
: LIBRARY: lambda gt11
: CLONE: KP23-1
: POSITION IN GENOME:
: UNITS: bp
: FEATURE:
: NAME/KEY: 5'UTR
: LOCATION: 1..324
: FEATURE:
: NAME/KEY: CDS
: LOCATION: 325..2511
: FEATURE:
: NAME/KEY: 3'UTR
: LOCATION: 2512..2675
: PUBLICATION INFORMATION:
: AUTHORS: Poulin, Matthew L
: TITLE: Nucleotide sequences of two new
: TITLE: factor receptor-2 variants
: JOURNAL: Biochim. Biophys. Acta
: VOLUME: 1220
: PAGES: 209-211
: DATE: 1994
: RELEVANT RESIDUES IN SEQ ID NO: 5: FROM 1 TO 2675
: US-08-885-418-5

alignment_scores:
: Quality: 75.00 Length: 202
: Ratio: 0.833 Gaps: 10
: Percent Similarity: 44.354 Percent Identity: 22.277

alignment_block:
: US-09-151-189-2_COPY_441_676 x US-08-885-418-5
: Align seg 1/1 to: US-08-885-418-5 from: 1 to: 2675
:
: 44 LeuAspAlaaspPheaspIleaserLeuGluAsnAspGluLeuLeu 60
: .....: .....: .....: .....: .....: .....:
: 1336 GTGTCTGCTGCTCCAGCTCTCTATGATCCACACCTCCAGCTGGTGG 1385
:
: 60 sarGValAlaGluIleAsnAlaAlaGlnAsnProAsnAspGluValThr 77
: .....: .....: .....: .....: .....: .....:
: 1386 G.....ATCACACTCCCTGCTCTCCACATGAC...ACCC 1420
:
: 77 yrLeuLeuProGlnAlaIleGlnValGlySerProThrHisPro..... 91
: .....: .....: .....: .....: .....: .....:
: 1421 ACTGTGTGGCGGGGTCTCCAGTATGATGCTGCCAGAGCCGCCAAGTGG 1470
: .....: .....: .....: .....: .....: .....:
: 92 serTyrProSer.....GlyHisAlaThrGlnAsnGlyAl 103
: .....: .....: .....: .....: .....: .....:
: 1471 GAGTATCCAGGAGAAAGCTCAGCTGGGAGACCCCTGGCGAGGCTG 1520
: .....: .....: .....: .....: .....: .....:
: 103 aPheAlaThrValLeuLysAla.....LeuIleGlyLeuAspArgGlyGly 119
: .....: .....: .....: .....: .....: .....:
: 1521 CTCGCGGAGGTGGTGGTGGCAGAGCGGTGGCAGACAGAGCGGC 1570
: .....: .....: .....: .....: .....: .....:
: 119 lucyspPheProasnProVal.....PheProSerAspaspIleu 132

```

```

: 1571 CCAAGATGCAGCAGCAGCTGGCAGTCAAGATGCTGAAGACAGATGCAACC 1620
: .....: .....: .....: .....: .....: .....:
: 133 Glu.....: .....: .....: .....: .....: .....:
: 1621 GAGAGGATCTTTCTGATCTGCTGCTGAGATGGAAATGATGAAGATGAT 1670
: .....: .....: .....: .....: .....: .....:
: 134 .....LeuIleAsnPheGluGlyAlaCysLeuThrTyrG 145
: .....: .....: .....: .....: .....: .....:
: 1671 TGGGAAGCATAAATAATATCATCATCTTCTAGGAGCGTGC...ACCCAAG 1717
: .....: .....: .....: .....: .....: .....:
: 145 lucGlyGluIleAsnLysLeuAlaValAsnValAlaPheGly..... 158
: .....: .....: .....: .....: .....: .....:
: 1718 ATGCCCATCTCTACGTGATATGCTCGATATGCCCTCCAAAGGGAACTTGGT 1767
: .....: .....: .....: .....: .....: .....:
: 159 .....ArgGlnMetLeuGlyIleHisTyrArgPheAspG 170
: .....: .....: .....: .....: .....: .....:
: 1768 GAATACTTGGCAGCCCGCCGCCACCTGGCATGGAGTACTCTTTGAC... 1815
: .....: .....: .....: .....: .....: .....:
: 170 yileGlnGlyLeuLeuLeuGlyGluThrIleThrValArgThrLeuHisG 187
: .....: .....: .....: .....: .....: .....:
: 1816 .....ATCACACATATTCCTG 1831
: .....: .....: .....: .....: .....: .....:
: 187 lngIleuLeuMetThrPheAlaGlyGluAlaThrPheGluPheArgLeuPhe 203
: .....: .....: .....: .....: .....: .....:
: 1832 AAGAGCAGATGACCTTCAAGGACCTAGTGTCTTGCCAGCTACCACTGGCC 1881
: .....: .....: .....: .....: .....: .....:
: 204 ThrGly 205
: .....: .....: .....: .....: .....: .....:
: 1882 AGGGGA 1887

```

seq_name: /cgn2_6/ptodata/ina/5B_COMB.seq:US-08-070-165F-9

```

seq_documentation_block:
: Sequence 9, Application US/08070165F
: Patent No. 5750365
: GENERAL INFORMATION:
: APPLICANT: Chiu, Ing-Ming
: TITLE OF INVENTION: Acidic Fibroblast Growth Factor (aFGF)
: NUMBER OF SEQUENCES: 12
: CORRESPONDENCE ADDRESS:
: ADDRESS: Ing-Ming Chiu
: STREET: S2052 Davis Medical Research Center, 480 West
: STREET: 9th Avenue
: CITY: Columbus
: STATE: Ohio
: COUNTRY: USA
: ZIP: 43210
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/070,165F
: FILING DATE:
: CLASSIFICATION: 435
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (614)-293-8093
: TELEFAX: (614)-293-5631
: INFORMATION FOR SEQ ID NO: 9:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 2681 base pairs
: TYPE: nucleic acid
: STRANDEDNESS: double
: TOPOLOGY: linear
: MOLECULE TYPE: cDNA
: HYPOTHETICAL: NO
: ANTI-SENSE: NO
: ORIGINAL SOURCE:
: ORGANISM: No. 5750365ophthalmus viridescens
: DEVELOPMENTAL STAGE: Adult

```



```

153 TTTCTGAGCTCTACGCGAGCACTTGTACAGCTGGCGCGCAAGCTCTGC 202
      |||||
224 SerGlyLeuValThrGlyValAlaAspCys 234          29-SEP-1999
      |||||
203 GAGGGG...GCCTTACCGCGCAGCACTTGT 232
      |||||

seq_name: gb_est35.AL045228

seq_documentation_block:
LOCUS AL045228 468 bp mRNA EST
DEFINITION DKFDP434B2050_r1_434 (synonym: htes3) Homo sapiens cDNA clone
            DKFDP434H2050_5', mRNA sequence.
ACCESSION AL045228
VERSION AL045228.1 GI:5433390
KEYWORDS EST.
SOURCE human.
ORGANISM Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
            Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 468)
AUTHORS Poustka,A., Klein,N., Mewes,H.W., Gassenhuber,J. and Wiemann,S.
TITLE EST (Poustka, et al.)
JOURNAL Unpublished (1999)
COMMENT On May 1, 1997 this sequence version replaced gi:2059563.
            Contact: Poustka A.J.
            Department Lehrbach
            Max-Planck-Institute for Molecular Genetics
            Ihnestr. 73, 14195 Berlin, Germany
            Tel.: +49-30-84131623
            Fax: +49-30-84131128
            Email: Poustka@mpg-berlin-dahlem.mpg.de
            This is the 5' sequence of the clone insert
            Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
            Research Center (DKFZ); Email: s.wiemann@dkfz-heidelberg.de;
            Sequenced by DKFZ within the cDNA sequencing consortium of the
            German Genome Project.
            No 5' sequence available.
            This clone is available at the RZPD in Berlin.
            Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
            Berlin-Charlottenburg, GERMANY; Email: clone@rzpd.de.
FEATURES
            Location/Qualifiers
            1..468
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="DKF2P434H2050"
            /clone_lib="434 (synonym: htes3)"
            /tissue_type="testis"
            /dev_stage="adult"
            /lab_host="DH10B"
            /note="Vector: pSport1; Site_1: NotI; Site_2: SalI"
BASE COUNT 98 a 137 c 156 g
ORIGIN

```

```

alignment_scores:
Quality: 86.50 Length: 149
Ratio: 1.153 Gaps: 9
Percent Similarity: 50.336 Percent Identity: 28.859

alignment_block:
US-09-151-189-2_COPY_441_676 x AL045228
Align seg 1/1 to: AL045228 from: 1 to: 468

```

```

1 HistyPheArgLeuIleGlyAlaAlaGluLeuAlaGlnArgAlaSerCy 17
|||||
29 CACTACAGCAAGTC.....TCTCCAGAGCGACGCTGTCGGGAG.....67
17 styGlnIlys.....TpgGlnValHisArgPheAlaArgProGluAla 32
|||||
68 .....CAGAGCGCACTGTGGAGATCCAGG.....94

```

```

32 euGlyGlyThrLeuHisAsnThrIleAlaGlyAspLeuAspAlaAspPhe 48
      |||||
95 .....CACCAGCTGGAGAGTGCACGGCTTGT.....121
      |||||
49 AspIleSerLeuLeuGluAsnAsp.....GluLeuLeuIysArg 61
      |||||
122 .....TTTCTCTTAGAGATGAGAAACACATTCAGGAGATGCTGAAGC 165
      |||||
61 gValAlaGluIleAsnAlaAlaGlnAsnProAsnAsnGluValThrTyrL 78
      |||||
166 AGTCTCTGAGGCTCACCAGCCCGAGAGCCGCCAGAGCTCTCTTCAGGG 215
      |||||
78 euLeuProGlnAlaIleGlnValGlySerProThrHisProSerTyrPro 94
      |||||
216 TCCACCCCGAGGTCAGCTCTCGGGC...CCTGAGCCACCGAG.....256
      |||||
95 SerGlyHisAlaThrGlnAsnGlyAlaPheAlaThrValLeuIysAlaLe 111
      |||||
257 GGAGGCCAAGCGCTGATCTGGGCTCTCAGAGCAACACTGCGCTCCCTG 306
      |||||
111 uIleGlyLeuAspArgGlyGlyLucysPheProAsnProValPhePro. 127
      |||||
307 CGATGGCTGGAGAACGCCAGCGCTCTGACCCACACAGCCCTTAAAGCTC 356
      |||||
128 .....SerAspAspGlyLeuGluLeuLeuAsnPheGluGlyAla 140
      |||||
357 TAGCAGTGGGGGCTGTGTCAGGGGCTGCAGAGGTGGAGGGGGCC 403
      |||||

seq_name: gb_est47.AW401290

seq_documentation_block:
LOCUS AW401290 289 bp mRNA EST
DEFINITION LAM1Gest491est.L.digitata gametophyte Lambda ZapII Laminaria
            digitata cDNA similar to vanadium bromoperoxidase, mRNA sequence.
ACCESSION AW401290
VERSION AW401290.1 GI:6919794
KEYWORDS EST.
SOURCE Laminaria digitata
ORGANISM Laminaria digitata
            Eukaryota; Stramenopiles; Phaeophyceae/Xanthophyceae group:
            Phaeophyceae; Laminariales; Laminariaceae; Laminaria.
            1 (bases 1 to 289)
            Crepinneau,F., Roscoe,T., Kaas,R., Kloareg,B. and Boyen,C.
            Characterisation of complementary DNAs from the Expressed Sequence
            Tag analysis of life cycle stages of Laminaria digitata
            (Phaeophyceae)
            Unpublished (2000)
            On Dec 20, 1995 this sequence version replaced gi:1135359.
            Other ESTs: Lam1Gest491r7est
            Contact: Boyen C
            Centre d'Etudes Oceanologique et de Biologie Marine
            CNRS-LDR9042 Université P. & M. Curie
            BP74, P-29682 Roscoff cedex, France
            Tel: 33 2 98 29 23 32
            Fax: 33 2 98 29 23 24
            Email: est@sb-roscoff.fr
            Location/Qualifiers
            1..289
            /organism="Laminaria digitata"
            /db_xref="taxon:80365"
            /clone_lib="L.digitata gametophyte Lambda ZapII"
            /dev_stage="gametophyte"
            /lab_host="SOLR"
            /note="Vector: pBluescript SK+; Site_1: EcoRI; Site_2:
            XhoI; The Laminaria gametophyte library, constructed by T
            Roscoe and F Crepinneau, was oligo-(dr) primed and
            directionally cloned into a Uni-ZAPTM XR vector
            (Stratagene, La Jolla, CA, USA) using total mRNA from
            gametophytes provided by R Kaas, IFREMER Nantes, France."
BASE COUNT 71 a 88 c 77 g 53 t
ORIGIN

```


alignment_scores:
Quality: 83.50 Length: 126
Ratio: 1.246 Gaps: 9
Percent Similarity: 53.175 Percent Identity: 29.365

alignment_block:

US-09-151-189-2_COPY_441_676 x AI316774/rev

Align seg 1/1 to reverse of: AI316774 from: 1 to: 480

```

65 ILeasAlaAlaGlnAsnProAsnGlu.ValThrThrLeuLeuProG 81
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
381 CTTCATGTTATCATCTCAACCAATCCAGATCCCTCATGACACATCGCGG 332
81 InAlaIleGlnValGlySerProThrHISpProSerTyProSerGly... 96
331 GA.....ACAGCCCTGCCCTGCCATGCTTCAGGGTGG 294
97 .....HisAlaThrGlnAsnGlyAlaPheAlaThrValLeuLy 109
:|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
293 GAGGACCATATACAGCAGGAGAACTCTGCTGCCATATGCGTCTGTCAG 244
109 sAla.....LeuIleGly..... 113
243 GACAGGCGCTCCACGCCACCGCTCTCTAGGAGCTCCATATGCTC 194
114 ..LeuAspArgGlyGlyGlyCysPheProAsnProValPheProSerAsp 129
193 TTCTTACCGCGCGCGGCGAAGCAGCTCG..... 165
130 AspGlyLeuLeuLeuLeuLeuPheGlnCysLeuThrTyGlu... 145
164 CAGGCTCGAACTCGCGAGCAGAG.....GTCTTACTACGCGCGC 121
146 .....GlyGlyLeuLeuLeuLeuLeuLeu.....V 155
120 CTCCTGGCGGCGTGGAGCGCTGCAGCACTGCTCCATAGCCGAGCGCTCA 71
155 alAlaPheGlyArgGlnMetLeuGly 163
70 GCTCCACAGCGCGCGCGAGATGCTCGG 45

```

seq_name: gb_est24:AI227820

```

seq_documentation_block:
LOCUS AI227820 673 bp mRNA EST 20-JAN-1999
DEFINITION EST224515 Normalized rat brain, Bento Soares Rattus sp. cDNA clone
VERSION AI227820
KEYWORDS EST
SOURCE Rattus sp.
ORGANISM Rattus sp.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
REFERENCE
1 (bases 1 to 673)
Lee N.H., Glodex A., Chandra, I., Mason, T.M., Queckenbush, J.,
Kerlavage, A.R. and Adams, M.D.
Rat Genome Project: Generation of a Rat EST (EST) Catalog & Rat
Gene Index
Unpublished (1998)
Other ESTs: TC63251
Contact: Lee, NH
ATCC
The Institute for Genomic Research
9712, Medical Center Drive, Rockville, MD 20850, USA
Tel: (301)-838-3528
Fax: (301)-838-0208
Email: nhlee@igr.org
Seq primer: M13-21.
Location/Qualifiers
1..673
/organism="Rattus sp."

```

FEATURES

source

/db_xref="ATCC (inhost):2036189"

/db_xref="taxon:10118"

/clone="NR00011"

/clone_lib="Normalized rat brain, Bento Soares"

/note="Organ: brain; Vector: pT73pac; Site 1: EcoRI;

Site 2: NotI"

BASE COUNT 150 a 190 c 181 g 151 t 1 others

ORIGIN

alignment_scores:

Quality: 83.00 Length: 147
Ratio: 1.064 Gaps: 6
Percent Similarity: 53.061 Percent Identity: 23.810

alignment_block:

US-09-151-189-2_COPY_441_676 x AI227820/rev

Align seg 1/1 to reverse of: AI227820 from: 1 to: 673

```

34 GlyThrLeuHisAsnThrIleAlaGlyAspLeuAspAlaAspPheAsp11 50
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
485 GGCACCCGCCATGGA...GTGCCCGAGGAGGAGGAGGAGGAGGAGGAGT 439
50 sSerLeuLeuGlnAsnAspGluLeuLeuLeuValAlaGluIleAsnA 67
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
438 AGTCCTTGGACGGTTCACGTGCTGCAACAATCTCTGACAAAAACT 389
67 LeAlaGlnAsnProAsnGluValThrTyLeuLeuProGlnAlaIle 83
388 TGGAGTCCGATACCATCATCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 339
84 GlnValGlySerProThrHisProSerTyProSerGlyHisAlaThr100
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
338 ATGGTCATGGCAACACACGACGATGTCAGCTCTTCAACATCTGAGCCA 289
100 nasGlyAlaPheAlaThrValLeuLeuLeuValAlaLeuIleGlyLeuAspArg 117
289 GCTGCTCATACGCGCCACAGT..... 267
117 lyGlyGlyCysPheProAsnProValPheProSerAspGlyLeuGlu 133
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
266 .....TTCCCGACGCTCATGCGGACGACGAGATGAGGCGCTT... 231
134 LeuLeuAsnPheGlyGlyAlaCysLeuThrTyGlyGlyGlyLeuIleAsnLy 150
230 .....GGGTGTCACACT.....GG 217
150 sLeuAlaValAsnValAlaPheGlyValGlnMetLeuGlyIleHisTyA 167
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
216 GCTTCCAGTACCTACA.....AAGAGGTACTGTCTCATC.....C 182
167 gPheAspGlyIleGlnGlyLeuLeuLeuGlyGlyThrIle 180
|||||:|||||:|||||:|||||:|||||:|||||:|||||:|||||
181 GCATTGACCGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 141

```

seq_name: gb_est23:AI151044

seq_documentation_block:

```

LOCUS AI151044 470 bp mRNA EST 26-OCT-1998
DEFINITION qb63q01.x1 NCI-CGAP Brn23 Homo sapiens cDNA clone IMAGE:1704816 3'
similar to SW:NUBM_BOVIN_P25708 NADH-UBIQUINONE OXIDOREDUCTASE 51
XD SUBUNIT PRECURSOR,, mRNA sequence.

```

ACCESSION AI151044

VERSION AI151044.1 GI:3679513

KEYWORDS

EST.

ORGANISM

human.

REFERENCE

1 (bases 1 to 470)

NCI/NINDS-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.

National Cancer Institute / National Institute of Neurological

466 CTAATAACACAGAGCTGGCTTATCATCATAGCTGGAGCTAGGAA 417

73 AsnGluValThrTyrLeuLeuProGlnAlaIleGlnValGlySerPro 89

416 CGGAAGCTGCACCTCTTTCTGTAGCATCACTGGTCACATCTGGCGGCG 367

89 z.....HisProSerTyrProSerGlyHisAlaThrGlnAsnGlyAlaP 104

366 TCTTGGACATAAAGGGTACTCTGCGGCTCATAGCTTGGGAATACCCACT. 318

104 heAlaThrValLeuLysAlaLeuIleGlyLeuAspArgGlyGlyLeuCys 120

317CGAAACAAATCTGGAATCT 297

121 PheProAsnProValPheProSerAspGlyLeuGluLeu 134

296 CTTTGGACACCATCAACACCGGTCAACATCGGGTGCATCTA 255

```

seq_documentation_block:
LOCUS       796 bp                mrna                09-DEC-1999
DEFINITION  mgae005ba01f Magnaporthe grisea Appressorium stage cDNA library
            Pyricularia grisea cDNA clone mgae005ba01f 5', mRNA sequence.
ACCESSION   AI069032
VERSION     AI069032.1  GI:3392007
KEYWORDS    EST.
SOURCE      Pyricularia grisea
            Pyricularia grisea
ORGANISM    Pyricularia grisea; Ascomycota; anamorphic Ascomycota; Pyricularia.
REFERENCE   1 (bases 1 to 796)
AUTHORS     Chou, W., Fang, E., Sasnowski, M., Wing, R. and Bean, R.A.
TITLE       Expressed sequence characterization during appressorium formation
            in rice blast fungus, Magnaporthe grisea
UNPUBLISHED (1998)
COMMENT     On Jan 19, 1998 this sequence version replaced gi:2287141.
JOURNAL
COMMENT

```

FEATURES
source
1. 796
Location/Qualifiers
High quality sequence stop: 364.
Seq primer: T3 primer (AATTAACCTCACTAAGGG)
Email: rdeanclemson.edu
Fax: 864 656 4293
Tel: 864 656 5737

```

/cclone_mga6005DA011*
/cclone_lib+MagnaPorthe grisea Appressorium Stage cDNA
library
/dev_stage+Germinated conidia on appressorium-inductive
surface
/note+vector: pBluescriptII SK(+) vector; site.1: EcoRI
site.2: XhoI; The appressorium formation-specific cDNA
library was constructed from conidia germinated for 5-8
hr on an inductive surface. The library has an average
insert size of 1.5 kbp.
168 a 227 c 231 g 170 t
BASE COUNT-
ORIGIN

```

```

7  GlyAlaAlaGluLeuAlaGlnThrGalaSerCys...TyrGlnLysTrpEl 22
   |||||
68  GGGGCGCTCGCGCCCTGCTGGCGCCGCCCACTTAAGACTGCCCAAACTC 117
   |||||

22  nVal...HisArgPheAlaArgProGluAlaLeuGly...GlyThrLeuH 37
   |||||
118  TCTCGGTGCGAGGTTCGACGACACAGAGGTCTCGGTGATGGAAGATTC 167
   |||||

37  IAsnThrIleAlaGlyAspLeuAspAlaAspPheAspIleSerLeuLeu 53
   |||||
168  ACCAGCTCATTTGGTGGCGTCTGCTGGATGTGAAATTCGACACTGCC... 211
   |||||

54  GluAsnAspGluLeuLeuLeuAlaGValAlaGluIleAsnAlaAlaGlnAs 70
   |||||
212  ...AAGTCCCGCCCGATCTTCAACGCCCTTGAGAC 243

70  nProAsnAsnGluValThrTyrLeuLeuLeuProGluAlaIleGlnValGly 87
   |||||
244  CACCACACACACAGCAAGCTGCTCTGTGAG... 274

87  erProThrIleAspSerTyrProSerGlyHisAlaThrGlnAsnGlyAla 103
   |||||
275  ...GTTTCGCACATCTGCGCGAGAATGTCGC 304

104  PheAlaThrValLeuLeuAlaLeuIleGlyLeuAspArgGlyGlyLeuCy 120
   |||||
305  CGCTGCAATTCGATGCGAGGTACTAGGTGCTCACTGCTGCGCCCAAGCG 354
   |||||

120  sPheProAsnProValPheProSerAspAspGlyLeuLeuLeuIleAsnP 137
   |||||
355  C...AAGACACAGAGT... 367

137  heGluGlyAlaCysLeuThrTyrGluGlyGluIleAsnLysLeuAlaVal 153
   |||||
368  ...GCTCCCATC 376

154  AsnValAlaPheGlyAspGlnMetLeuGlyIleHisTyrArgPheAspGl 170
   |||||
377  ACCATTCGCGTGGCTCTGGAACCCCTGTGT...CGATCATGAA 417
   |||||

170  yIleGlnGlyLeuLeuLeuGlyGlnThrIleThrValaArgThr... 184
   |||||
418  CGTCACTGCTGACCCCATTCACAGCGGTGGTCCCATCAGACCGCAAGT 467
   |||||

185  ...LeuHisGlnGluLeuMetThrPheAlaGluGluAlaThrPhe 198
   |||||
468  ACCTTCCCATTCACGCTGAGGTGCCGAGATTCTGTCGAGCACTGCACC... 514
   |||||

199  GluPheArgLeuPheThrGlyGlu 206
   |||||
515  ...ACTGCGCAG 523

```

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT

Eutheria, Primates: Catarrhini, Hominoidea.
1 (bases 1 to 446).
HGCP <http://www.hgcp.org.br/ORESTES>.
THE PAPSEP/NIH Human Cancer Genome Project.
Unpublished (1999).
On Dec 20, 1995 this sequence version replaced g1:110000.
Contact: Simpson A.G.
Laboratory of Cancer Genetics
Ludwig Institute for Cancer Research

53 euGlusnAspGluLeuLeuysArgVal...AlaGluileAsnAlaAla 68
 353 TCCAAAGTCCACCAGATAGCTAGTACAGGCTGTGTGTGTGTGTATCC 304
 69 GlnAsnPro.....AsnAsnGluValThrTyLeuLeu.....ProG1 81
 303 ACAAACCTGAGCTAGACACAGAGTGTGTGTGTGTGTGTGTGTGTGTGT 254
 81 nAlaileGluValGlySerProThrHisProSerTyProSerGlyHisA 98
 253 AGCC.....AGCATPANAAGTTCTCCNAGTCCCACT 222
 98 laThrGlnAsnGlyAlaPheAlaThrValLeuLeuysAlaLeuileGlyLeu 114
 221 TGACTCAGGAGCCAGCTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGTGT 172
 115 AspArgGlyGlyGluCysPheProAsnProVal 125
 171 CCAGCGGGGACCCCTGCACCTCTCAGCCCTTG 139
 seq_name: gb_est38:AL120080

seq_documentation_block: 750 bp mRNA EST 27-SEP-1999
 LOCUS AL120080
 DEFINITION DKF2P761M212.1 761 (synonym: hamy2) Homo sapiens CDNA clone
 ACCESSION AL120080
 VERSION AL120080.1 GI:5925979
 KEYWORDS EST.
 SOURCE human.
 ORGANISM Homo sapiens
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Primates; Catarrhini; Hominidae; Homo.
 1 (bases 1 to 750)
 Ottenwaelder, B., Obermaier, B., Mewes, H.W., Gassenhuber, J., and
 Wiemann, S.
 EST (Ottenwaelder, et al.)
 Unpublished (1999)
 On Jun 22, 1998 this sequence version replaced gi:3247433.
 Contact: Ottenwaelder B
 MIPS
 Am Klopferplatz 18a D-82152 Martinsried, Germany
 This is the 5' sequence of the clone insert
 Clone from S. Wiemann, Molecular Genome Analysis, German Cancer
 Research Center (DKFZ); Email: s.wiemann@dkfz-heidelberg.de;
 sequenced by MedGenomix within the CDNA sequencing consortium of
 the German Genome Project.
 No 5' sequence available.
 This clone is available at the RZPD in Berlin.
 Please contact the RZPD: Ressourcenzentrum, Heubnerweg 6, 14059
 Berlin-Charlottenburg, GERMANY; Email: clonedrzd@de.
 Location/Qualifiers
 1..750
 /organism="Homo sapiens"
 /db_xref="taxon:9606"
 /clone_lib="761 (synonym: hamy2)"
 /tissue_type="amygdala"
 /dev_stage="adult"
 /lab_host="DH10B"
 /note="vector: pSport1; Site_1: NotI; Site_2: SalI"

BASE COUNT 173 a 181 c 213 g 183 t
 ORIGIN
 alignment_scores:
 Quality: 78.50 Length: 135
 Ratio: 1.172 Gaps: 7
 Percent Similarity: 49.630 Percent Identity: 28.148
 alignment_block:
 US-09-151-189-2_COPY_441_676 x AL120080/rev

Align seg 1/1 to reverse of: AL120080 from: 1 to: 750
 89 ThrHisProSer...TyProSerGlyHisAlaThrGlnAsnGlyAlaPhe 104
 386 ACACATGAGCCCTAATACAGGTGGGGAGAGAGCTCAAGAAAAGGTG 337
 104 eAlaThrValLeuysAlaLeuileGlyLeuAspArgGlyGlyGluCysP 121
 336 GGAGGGGATCTTTACTCACTCTCGTCCTGAT..... 303
 121 heProAsnProValPheProSerAspAspGlyLeuGluLeuileAsnPhe 137
 302ACCCAGCGACTCCCTCTGTGTCAGTCT 276
 138 GluGlyAlaCysLeuThrTyGluGlyGluileAsnLysLeuAlaValas 154
 275 GAGGTGCTCTGC.....GCCCTGCCCTGATC 250
 154 nValAlaPheGlyArgGlnMetLeuGlyLleHisTyArgPheAspGlyI 171
 249 CATAGCCTTTGGGCCCATACTCTGCGTA.....GCAGGAC 212
 171 leGlnGlyLeuLeuLeu.....GlyGluThrIleThrValArgThr 184
 211 TTGCAGTAATCTCTCCACCATGCAGCCACAGTGTGTGTGTGTGTGTGT 162
 185 Leu.....HisGlnGluLeuMetThrPheAlaGluLeuAlaThrPhe 198
 161 CTCTTCAGACCTTCACGACGACGACGAGGATTTATGAGAGCTGTGTCTT 112
 198 eGluPheArgLeuPheThrGlyGluValIleLysLeuPheGlnAspGlyT 215
 111 CGCATGAAGCTCTTC...GCCAAGTAACCGCTCTTCTGACACACACCCA 65
 215 hrPhe 216
 64 CATT 60

seq_name: gb_est24:AI232282
 seq_documentation_block: 618 bp mRNA EST 31-JAN-1999
 LOCUS AI232282
 DEFINITION EST228970 Normalized rat kidney, Bonto Soares Rattus sp. CDNA clone
 R18271 3' end, mRNA sequence.
 ACCESSION AI232282
 VERSION AI232282.1 GI:3816162
 KEYWORDS EST.
 SOURCE Rattus sp.
 ORGANISM Rattus sp.
 Eukaryota; Chordata; Craniata; Vertebrata; Mammalia;
 Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
 1 (bases 1 to 618)
 Lee, N.H., Glodek, A., Chandra, I., Mason, T.M., Quackenbush, J.,
 Kerlavage, A.R. and Adams, M.D.
 Rat Genome Project: Generation of a Rat EST (RSET) Catalog & Rat
 Gene Index
 Unpublished (1998)
 On Jan 14, 1998 this sequence version replaced gi:1797454.
 Other ESTs: TC62480
 Contact: Lee, NH
 ATCC
 The Institute for Genomic Research
 9712 Medical Center Drive, Rockville, MD 20850, USA
 Tel: (301)-838-3529
 Fax: (301)-838-0208
 Email: nhlee@tigr.org
 Seq primer: M13-21.
 Location/Qualifiers
 1..618
 /organism="Rattus sp."
 /db_xref="ATCC (lnhost):2037931"
 /db_xref="taxon:10118"
